PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/86, A61K 39/285, C12N 15/12, C07K 15/04, C12N 15/52 — A1

(11) International Publication Number: WO 90/12101

(43) International Publication Date: 18 October 1990 (18.10.90)

(21) International Application Number: PCT/GB90/00493

(22) International Filing Date: 3 April 1990 (03.04.90)

(30) Priority data:
8907468.6 3 April 1989 (03.04.89) GB

(71) Applicant (*for all designated States except US*): LYNXVALE LTD. [GB/GB]; The Old Schools, Cambridge CB2 1TS (GB).

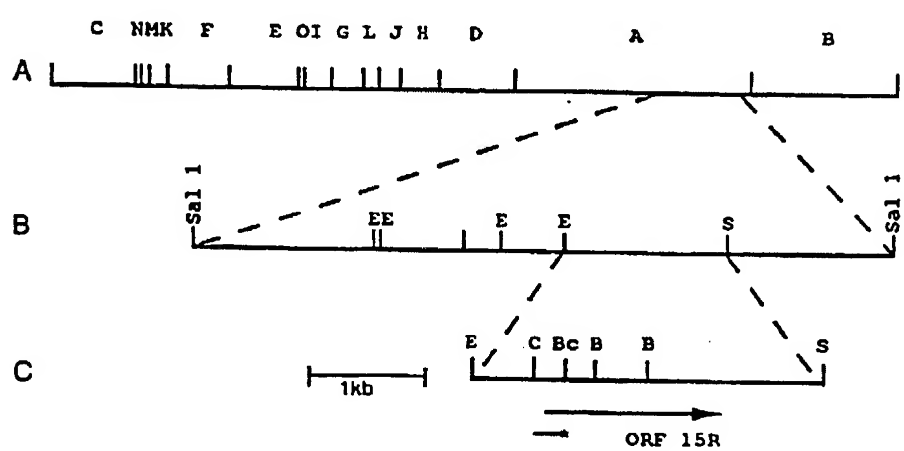
(72) Inventor; and
(75) Inventor/Applicant (*for US only*) : SMITH, Geoffrey, Lilley [GB/GB]; 44 Blandford Avenue, Oxford OX2 8DZ (GB).

(74) Agent: ARMITAGE, I., M.; Mewburn Ellis, 2 Cursitor Street, London EC4A 1BQ (GB).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB, GB (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), SE (European patent), US.

Published
With international search report.

(54) Title: VACCINIA VECTORS, VACCINIA GENES AND EXPRESSION PRODUCTS THEREOF



(57) Abstract

The invention discloses recombinant vaccinia virus vectors wherein: a) part or all of one or more of the following nucleotide sequences is deleted from the viral genome; and/or b) one or more of said nucleotide sequences is inactivated by mutation or the insertion of foreign DNA; and/or c) one or more of said nucleotide sequences is changed to alter the function of the protein product encoded by said nucleotide sequence; which nucleotide sequences are sequences designated herein as i) Sal F 3R, ii) Sal F 9R, iii) Sal F 13R, iv) B 5R, v) Sal F 15 R.

+ See back of page

DESIGNATIONS OF "DE"

Until further notice, any designation of "DE" in any international application whose international filing date is prior to October 3, 1990, shall have effect in the territory of the Federal Republic of Germany with the exception of the territory of the former German Democratic Republic.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MR	Mauritania
BE	Belgium	GA	Gabon	MW	Malawi
BF	Burkina Fasso	GB	United Kingdom	NL	Netherlands
BG	Bulgaria	HU	Hungary	NO	Norway
BJ	Benin	IT	Italy	RO	Romania
BR	Brazil	JP	Japan	SD	Sudan
CA	Canada	KP	Democratic People's Republic of Korea	SE	Sweden
CF	Central African Republic			SN	Senegal
CG	Congo	KR	Republic of Korea	SU	Soviet Union
CH	Switzerland	LI	Liechtenstein	TD	Chad
CM	Cameroon	LK	Sri Lanka	TG	Togo
DE	Germany, Federal Republic of	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

VACCINIA VECTORS, VACCINIA GENES AND EXPRESSION PRODUCTS THEREOF

BACKGROUND OF INVENTION

FIELD OF INVENTION

The present invention relates to recombinant vaccinia virus vectors. In particular it relates to the attenuation of the virus, to potential enhanced immunogenicity of the virus, to the provision of sites for the insertion of heterologous gene sequences into the virus, and to the use of the recombinant virus vectors thereby provided. It also relates to proteins which are the expression products of vaccinia genes.

DESCRIPTION OF PRIOR ART

Live vaccinia virus was used as the vaccine to immunise against, and eradicate smallpox. Vaccinia virus is the prototypical member of the poxvirus family and therefore it has been extensively studied. It is a large DNA-containing virus which replicates in the cytoplasm of the host cell. The linear double-stranded genome of approximately 185,000 base pairs has the potential to encode at least 200 proteins (Moss, B. (1985) In B.N. Fields, D.M. Knipe, J.L. Melnick, R.M. Channock, B.R. Roizman and R.E. Shope (eds.), Virology. Raven Press, New York, pp. 685-704). The cytoplasmic site of replication requires that vaccinia virus encodes many enzymes and protein factors necessary for DNA synthesis. Advances in molecular genetics have made possible the construction of recombinant vaccinia viruses that contain and express genes

derived from other organisms (for review see Mackett, M. & Smith G.L. (1986), J. Gen. Virol., 67, 2067-2082). The recombinant viruses retain their infectivity and express the foreign gene (or genes) during the normal replicative cycle of the virus. Immunisation of animals with the recombinant viruses has resulted in specific immune responses against the protein(s) expressed by the vaccinia virus, including those protein(s) expressed by the foreign gene(s) and in several cases has conferred protection against the pathogenic organism from which the foreign gene was derived.

Recombinant vaccinia viruses have, therefore, potential application as new live vaccines in human or veterinary medicine. Advantages of this type of new vaccine include the low cost of vaccine manufacture and administration (because the virus is self-replicating), the induction of both humeral and cell-mediated immune responses, the stability of the viral vaccine without refrigeration and the practicality of inserting multiple foreign genes from different organisms into vaccinia virus, to construct polyvalent vaccines effective against multiple pathogens. A disadvantage of this approach, is the re-use of a virus vaccine that has been recognised as causing rare vaccine-related complications.

The applicants have now identified unobvious gene sequences which may be deleted from the viral genome. The applicants propose that part or all of one or more of these gene sequences may be deleted from the viral genome to allow (i) greater attenuation of the virus; and/or (ii)

enhancement of immunogenicity of recombinant vaccinia virus; and/or (iii) further gene sequence insertion sites so that more foreign DNA may be included in the virus. Where however, the gene sequences are essential for viral replication, viral attenuation can still be effected by altering the gene product (e.g. by manipulation at gene level) such that a protein function affecting pathogenicity is adversely affected whilst keeping the protein functional for virus application.

SUMMARY OF INVENTION

According to one aspect of the present invention there is provided a vaccinia virus vector wherein a) part or all of one or more of the following nucleotide sequences is deleted from the viral genome; and/or b) one or more of said nucleotide sequences is inactivated by mutation or the insertion of foreign DNA; and/or c) one or more of said nucleotide sequences is changed to alter the function of a protein product encoded by said nucleotide sequence; which nucleotide sequences are sequences designated herein as i) Sal F 3R, ii) Sal F 9R, iii) Sal F 13R, iv) B5R, v) Sal F 15R.

DNA sequences encoding one or more heterologous polypeptides may be incorporated in the viral genome. The DNA sequences encoding the heterologous peptides may be inserted into one or more ligation sites created by the deletion or deletions from the viral genome.

The recombinant vaccinia viruses of the present invention have the potential for enhanced immunogenicity. This may result from either the deletion of vaccinia genes

which cause immunosuppression (e.g. the complement homologue and the human FcR for IgE) or by insertion of a gene which potentiates the immune response (e.g. expressing the authentic CD23 gene in vaccinia virus).

Therefore the present invention provides a vaccinia virus wherein a) part or all of one or more vaccinia nucleotide sequences causing immunosuppression are deleted from the viral genome; and/or b) one or more of said vaccinia nucleotide sequences causing immunosuppression is inactivated by mutation or the insertion of foreign DNA; and/or c) one or more of said vaccinia nucleotide sequences causing immunosuppression is changed to alter the function of a protein product encoded by said nucleotide sequence; which nucleotide sequences are sequences designated herein as i) Sal F 3R, ii) Sal F 9R, iii) Sal F 13R, iv) B5R, v) Sal F 15R.

In particular the vaccinia nucleotide sequence may be the sequence designated herein as Sal F 3R.

Where the vaccinia virus comprises a DNA sequence encoding a heterologous polypeptide which potentiates the immune response, the DNA sequence may encode CD23.

The recombinant vaccinia vectors of the present invention may be used as immunogens for the production of monoclonal and polyclonal antibodies or T-cells with specificity for heterologous peptides encoded by DNA sequences ligated into the viral genome. The invention also provides the monoclonal antibodies, polyclonal antibodies, antisera and/or T cells obtained by use of the recombinant vaccinia vectors provided. The antibodies

produced by use of the recombinant virus vectors hereof can be used in diagnostic tests and procedures, for example in detecting the antigen in a clinical sample; and they can also be used therapeutically or prophylactically for administration by way of passive immunisation. Also provided are diagnostic test kits comprising monoclonal antibodies, polyclonal antibodies, antisera and/or T cells obtained by use of the recombinant vaccinia vectors provided.

Also provided are vaccines and medicaments which comprise a recombinant vaccinia virus hereof. These may have enhanced safety and immunogenicity over current vaccinia virus strains for the reasons indicated.

According to another aspect of the present invention there is provided a polypeptide encoded by a nucleotide sequence selected from those defined above and alleles and variants of said polypeptides. The polypeptide, allele or variant thereof may be encoded by the nucleotide sequence designated herein as Sal F 13R and which has activity as a DNA ligase.

The invention also includes sub-genomic DNA sequences encoding such a polypeptide, recombinant cloning and expression vectors containing such DNA, recombinant microorganisms and cell cultures capable of producing such a polypeptide.

The invention also provides a method of attenuating a vaccinia virus vector which comprises: a) deleting part or all of one or more of the following nucleotide sequences from the viral genome; and/or b) inactivating one or more

of said nucleotide sequences by mutating said nucleotide sequences or by inserting foreign DNA; and/or c) changing said one or more nucleotide sequences to alter the function of a protein product encoded by said nucleotide sequence; which nucleotide sequences are sequences designated herein as: i) Sal F 3R, ii) Sal F 9R, iii) Sal F 13R, iv) B5R, v) Sal F 15R.

The invention also provides a method which comprises using a vaccinia virus vector as defined herein to prepare a vaccine or a medicament.

The invention also provides the use of part or all of the nucleotide sequence designated herein as Sal F 13R or part or all of the amino acid sequence encoded by said nucleotide sequence in the identification of polypeptides with activity as a DNA ligase. Furthermore the polypeptide represented by the amino acid sequence 7625 to 9280 (inclusive) of Fig. 11 hereof, or an allele or variant thereof, may be used as an enzyme in the manipulation of DNA in recombinant technology.

BRIEF DESCRIPTION OF THE DRAWINGS

In order that the present invention may be understood more clearly, the identified gene sequences will be described more fully with reference to the Figures wherein:

Figure 1 shows the location and direction of transcription of DNA ligase gene within the vaccinia virus genome. A. Vaccinia virus HindIII restriction map. B. The 13.4 kb SalI F restriction fragment is expanded and the positions of EcoRI(E) SmaI(S) restriction sites are indicated. C.A 3300bp EcoRI - SmaI fragment is expanded

showing the position of the DNA ligase gene (arrow), and the positions of EcoRI(E), ClaI(C), BclI(Bc), BglII(B) and SmaI(S) restriction sites;

Figure 2 shows the location and direction of transcription of a thymidylate kinase gene within the vaccinia virus genome. A. Hind III restriction map. B. Expanded 13.4 kb SalI F fragment with the position of SalI F 13R (ORF13) shown as filled box. C. Expanded 2.4 kb Dra I fragment with position and direction of transcription of ORF13 shown. The scale refers to this fragment. Letters in B and C indicate restriction enzyme sites: EcoRI(E), DraI(D), BamHI (B) and BclI(BC).

Figure 3 shows the nucleotide sequence of a 1776 nucleotide region of the 13.4 kb SalI F fragment. The deduced sequence of a 552 amino acid open reading frame is shown. The open reading frame is designated Sal F 15R;

Figure 4 shows the nucleotide sequence of 800 bp region of the vaccinia virus SalI F fragment. The deduced amino acid sequence of a 227 amino acid open reading frame designated Sal F 13R is shown. Numbers on the upper and lower lines refer to amino acids from the beginning of the ORF or to the nucleotides from start of DNA fragment, respectively. Underlined nucleotides represent potential early transcriptional termination sequences and asterisks represent the 5' ends of early mRNA determined by S1 nuclease protection.

Figure 5 shows the nucleotide sequence and amino acid sequence for the gene Sal F 3R;

Figure 6 shows the amino acid sequence homology

between the protein encoded by gene Sal F 3R and i) the human low affinity Fc receptor for IgE, (huFcR(IgE)); ii) the antifreeze polypeptide (ANP) from Hemitripterus americans; and iii) a lectin (LEC) from Megabalanus rosa;

Figure 7 shows the construction of plasmid pSAD3G for the deletion of Sal F 3R from the virus genome;

Figure 8 shows a Southern blot analysis of virus vSAD3. Virus DNA was extracted from purified WT or vSAD3 virus and digested with SpeI. DNA fragments were resolved on an agarose gel, transferred to nitrocellulose and probed with a radio-labelled DNA fragment from the Ecogpt gene. The band of 7 kb is as predicted and there is no hybridization with DNA from WT virus;

Figure 9 shows a Northern blot of mRNA from mock-infected (lane 1) or WT virus-infected cells early (lane 3) or late (lane 2) after infection. RNAs were resolved on an agarose gel transferred to nitrocellulose and probed with a single stranded, radio-labelled DNA fragment complementary only to the Sal F 3R open reading frame;

Figure 10 shows the nucleotide and amino acid sequence for the gene Sal F 9R;

Figure 11 shows the amino acid sequence homology between the protein encoded by gene Sal F 9R (SalI F ORF9) and i) cow; and ii) human, superoxide dismutase (SOD) (Cu-Zn) proteins;

Figure 12 shows the amino acid sequence homology between the protein encoded by gene Sal F 13R and yeast thymidylate kinase (TmpK);

Figure 13 shows aligned amino acid sequences of

vaccinia virus Sal F 13R (VV) and Saccharomyces cerevisiae (SC) TmpK. Identical amino acid residues are boxed. Numbers above or below the aligned sequences refer to amino acid positions of VV or SC respectively;

Figure 14 shows : A. the aligned amino acid sequences for the presumed ATP binding site of vaccinia (VV) and Saccharomyces cerevisiae (SC) TmpK, HSV TK/TmpK and human and VV TK. Residues identical in all 5 sequences are boxed. Numbers indicate the amino acids between the amino terminus and the region shown; and B. Amino acid sequences for region of HSV TK/TmpK involved in nucleoside/nucleotide binding, aligned with corresponding regions of vaccinia virus (VV) or Saccharomyces cerevisiae (SC) TmpK proteins. Amino acids conserved between two or all, of the sequences are boxed.

Figure 15 shows the biochemical pathway of dTTP synthesis in which thymidylate kinase is active;

Figure 16 shows the construction of plasmids pACV1 and pACV2;

Figure 17 shows a Southern blot analyses of viruses vACHB and vAC1. Virus DNA was extracted from purified WT, vAC1 or vACHB viruses and digested with SalI. DNA fragments were resolved on an agarose gel, transferred to nitrocellulose and probed with a radio-labelled DNA fragment from entirely within the TmpK gene. SalI digest gives a 13.4kb band with WT virus but bands of 8.8 and 6.7 kb for recombinants VACHB and VAC1 (due to an extra SalI site introduced at 3' end of Ecogpt cassette);

Figure 18 shows the nucleotide and amino acid sequence

for the gene B5R;

Figure 19 shows the amino acid sequence homology between the protein encoded by gene B5R (SalI G ORF10) and i) coagulation factor XIII B chain (F13 B); ii) complement factor H precursor (CFAH); iii) complement C2 precursor (CO2); and iv) complement C4B-binding protein precursor (C4BP);

Figure 20 shows the hydrophobicity profiles for B5R (SalI G ORF10) and H3C 28K proteins;

Figure 21 shows Northern blot of mRNA from virus infected cells early (E) or late (L) during infection. RNAs were resolved on an agarose gel transferred to nitrocellulose and probed with a single stranded, radio-labelled DNA fragment complementary only to the B5R gene. The position of molecular weight size markers is shown in kb;

Figure 22 shows the amino acid sequence homology between the vaccinia virus (VV) protein encoded by gene Sal F 15R and amino acid sequences of yeast DNA ligases from <u>S.pombe</u> (sp) and <u>S.cerevisiae</u> (sc) made using programme MULTALIGN;

Figure 23 shows the identification of vaccinia virus DNA ligase protein. Crude extracts were prepared from mock infected or vaccinia virus infected (100 pfu/cell) CV1 cells by Dounce homogenisation in 100 mM NaCl buffer as described in Kerr and Smith. Vaccinia virus infected early (lane 2), late (lane 3) or mock infected (lane 1) CV1 cell extracts and purified calf thymus DNA ligase I (a gift from T. Lindahl) (lane 5) were incubated with α-(^{32}P) ATP

(Methods). Reactions were terminated by trichloroacetic acid and covalently labelled polypeptides analysed by SDS PAGE on a 12.5% gel;

Figure 24 shows the 61 kD polypeptide is a DNA ligase. A DNA ligase preparation partially purified from vaccinia virus infected cells late (15h) post infection was labelled with α-(^{32}P) ATP (lane 3). Preparations of calf thymus DNA ligase (a gift from T.Lindahl, ICRF) (lane 1) and bacteriophage T4 DNA ligase (New England Biolabs) (lane 2) were labelled in parallel. The vaccinia sample was divided into four equal parts. One part was analysed without further manipulation (lane 3) and the remainder centrifuged through a column to remove unincorporated ATP as described in (25) except that Sephadex-G25 was used. The excluded volume was divided into three equal parts and incubated at 37°C for 30 minutes with either no addition (lane 4), cold poly (dA):oligo (dT) DNA ligase substrate (lane 5) or 100µM sodium pyrophosphate (lane 6). The products were analysed as in Figure 1;

Figure 25 shows DNA ligase activity in vaccinia virus infected cells. Crude extracts from CV-1 cells infected with vaccinia virus early (3h), late (17h) post infection or mock infected were assayed for DNA ligase activity (Kerr and Smith, Nucleic Acids Res., 17, 9039 (1989)). The 30mer (^{32}P) oligo dT:poly dA substrate is shown in lane 1 and corresponds to the monomer n = 1. Four units (lane 2), 0.4 units (lane 3) and 0.04 units (lane 4) of bacteriophage T4 DNA ligase (New England Biolabs) were assayed in parallel and provide markers (n = 2, n = 3, n = 4). Lanes

5, 6 and 7 represent the supernatant fractions from early, mock and late samples respectively after Dounce homogenisation and centrifugation at 10K for 20 minutes. Lanes 8, 9 and 10 are the pellet fractions from early, mock and late samples. An autoradiograph of the dried gel is shown;

Figure 26 shows: A. Immune-precipitation of (^{35}S)-methionine labelled polypeptides from vaccinia virus infected cells. TK cells infected with vaccinia virus (30 pfu/cell) or mock infected were labelled with (^{35}S)-methionine 1.5 - 4h post infection. Cell extracts were prepared ad immune-precipitated with pEX LIG antiserum (Kerr and Smith 1989). Lane 1 represents uninfected cells and lane 2 vaccinia virus infected cells. Molecular weight markers are shown to the right of the gel; and B. Co-migration of (^{35}S)-methionine and α-(^{32}P)-ATP labelled proteins. A (^{32}P)-labelled DNA ligase-AMP adduct from vaccinia virus infected cells (lane 1) and cell extracts labelled with (^{35}S)-methionine 2.5 - 6h p.i. from either vaccinia virus infected (lane 2) or mock infected cells (lane 3), immune-precipitated with pEX LIG antiserum as described in Part A, were electrophoresed through a 12.5% polyacrylamide gel;

Figure 27 shows immune-precipitation of labelled vaccinia virus DNA ligase. Calf thymus DNA ligase (lane 1), T4 DNA ligase (lane 2) and a phosphocellulose column fraction from vaccinia virus infected cells (lane 3) were incubated with α-(^{32}P)-ATP (Kerr and Smith 1989). Each sample was divided into four equal parts and either

analysed directly by TCA precipitation and SDS-PAGE (lanes 1, 2 and 3) or, in the case of extract from vaccinia virus infected cells, immune-precipitated with either pre-immune serum (lane 4), pEX LIG serum (lane 5) or a non-specific pEX immune serum (lane 6), followed by SDS-PAGE;

Figure 28 shows the cloning of the vaccinia virus DNA ligase gene by PCR to form plasmid pSK17;

Figure 29 (i) shows the expression of the DNA ligase gene in <u>E.coli</u>. Bacteria harbouring either parent vector pGMT7 or plasmid pSK18, were incubated with (+) or without (-) IPTG and the total cell protein run on an SS-polyacrylamide gel 4 hours after induction. The presence of an additional band of roughly 63 kDa is evident in bacteria containing pSK18 after addition of IPTG; (ii) timecourse of induction of DNA ligase after addition of IPTG. Bacterial cultures containing plasmid psK18 were inbucated for 0, 1, 2 or 4 hours after addition of IPTG and the total bacterial protein run on a polyacrylamide gel. The DNA ligase protein appears as a 63 kDa protein; (iii) the cell extracts shown in Figure 4b(i) were incubated with alpha-labelled ^{32}P-ATP and run on a polyacrylamide gel and an autoradiograph produced. The DNA ligase binds AMP and appears as a 63 kDa protein;

Figure 30 shows a Southern blot of virus DNAs from viruses derived from cells infected with WT vaccinia virus and transfected with pSK14. DNA was digested with SalI, run on an agarose gel and probed with the region of the DNA ligase gene deleted from pSK14. Isolates 3, 6, 7 and 8 lack the DNA ligase sequence but replicate efficiently in

tissue culture; and

Figure 31 shows covalent binding of alpha-labelled ^{32}P-ATP to extracts of cells infected with viruses 1, 5, 7 and 8 described in Figure 30. Viruses 7 and 8 lack a DNA ligase protein consistent with the lack of DNA for the gene product shown in Figure 30.

DESCRIPTION OF EMBODIMENTS

All the genetic manipulations described below were carried out according to standard procedures (Molecular Cloning, eds. Sambrook, Fritsch & Maniatis, Cold Spring Harbor Laboratory Press, 1989) and the conditions used for enzymatic reactions were as recommended by the manufacturer (GIBCO-BRL Life Technologies).

Determination of the Nucleotide and Amino Acid Sequences.

The nucleotide sequence of the SalI F and SalI G restriction fragments of the vaccinia virus genome (strain WR) were determined by established methods (Sanger, F. et al. (1980), J. Mol. Biol., 143, 161-178) and Bankier, A. and Barrell, B.G. (1983) Techniques in Life Sciences B508., 1-34, Elsevier.

For example, the 13.4 kb SalI F fragment of vaccinia virus (strain WR) was isolated from cosmid 6, which contains virus DNA derived from a rifampicin resistant mutant (Baldick, C.J. & Moss, B. (1987) Virology 156, 138-145), and was cloned into SalI cut pUC13 to form plasmid pSalI F. The SalI fragment was separated from plasmid sequences and self-ligated with T4 DNA ligase. Circular molecules were randomly sheared by sonication, end-repaired with T4 DNA polymerase and Klenow enzyme and fragments of

greater than 300 nucleotides cloned into SmaI cut M13mp18. Single stranded DNA was prepared and sequenced using the dideoxynucleotide chain termination method (Sanger, F., Nicklen, S. & Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463-5467), using [^{35}S]-dATP and buffer gradient polyacrylamide gels (Biggin, M.D., Gibson, T.J. & Hong, G.F. (1983), Proc. Natl. Acad. Sci. USA, 80, 3693-3695). For further details see (Bankier, A.T., Western, K.M. & Barrell, B.G. (1987) in Wu R. (ed.) Methods in Enzymology 155, 51-93. Academic Press, London). The 12.6 SalI G fragment was similarly treated.

Computer analysis

Nucleotide sequence data were read from autoradiographs by sonic digitiser and assembled into contiguous sequences using programmes DBAUTO and DBUTIL (Staden, R. (1980) Nucleic Acids Res. 8, 3673-3694; Staden, R. (1982) Nucleic Acids Res. 10, 4731-4751) on a VAX 8350 computer. The consensus sequence was translated in 6 frames using programmes ORFFILE and DELIB (M. Boursnell, Institute of Animal Health, Houghton, UK.). Open reading frames were compared against SWISSPROT protein database and against the applicants own database of vaccinia amino acid sequences using programme FASTP (Lipman, D.J. & Pearson, W.R. (1985) Science 227, 1435-1441). Alignments of multiple protein sequences were performed using programme MULTALIGN (Barton, G.J. & Sternberg, M.J.E. (1987) J. Mol. Biol. 198, 327-337).

There follows a description of the individual gene sequences the applicants have identified. Genes are named

by (i) the restriction fragment from which they derive (i.e. SalF means SalI F fragment, or B means HindIII B fragment), (ii) the number of the open reading frame initiating within the restriction fragment starting from the left end and (iii) the direction of transcription leftwards (L) or rightwards (R).

1. Sal F 3R

The nucleotide sequence and deduced amino acid sequence of the gene designated Sal F 3R is shown in Figure 5. The single letter code is used for the designation of amino acids. The coding region of the gene maps between nucleotides 595 and 1071 from the left end of the SalI F fragment. The molecular weight of the primary translation product is predicted to be 18.1 kiloDaltons (kD). Near the amino terminus there is a string of hydrophobic amino acids thought to cause the protein to be associated with, or secreted through, the cell membrane. Near the carboxy terminus there are three potential N-linked glycosylation sites, indicating that the mature gene product is a glycoprotein.

Comparisons of the deduced amino acid sequence with the protein database SWISSPROT established several significant homologies. Three of these are shown in Figure 6. The amino acid sequence encoded by the gene Sal F 3R shows sequence homology to a variety of lectins and the nearest homologue is human CD23 (see later). In particular, the amino acid sequence encoded by the gene Sal F 3R shows sequence homology with the amino acid

sequence of the human low affinity Fc receptor for IgE (Kitutani, H. et al. (1986), Cell, 47, 657), the amino acid sequence of an antifreeze polypeptide from Hemitripterus americans (see Ng, N.F. et al, (1986), J. Biol. Chem., 261, 15690-5) (5) and the amino acid sequence of a lectin from Megabalanus rosa (acorn barnacle) (Maramoto, K. & Kamiya, H. (1986), Biochem. Biophys. Acta., 874, 285-295.). Sal F 3R has a 26.1% amino acid identity over a 92 amino acid region of the human low affinity Fc receptor (FcR) for IgE, 22.4% amino acid identity over a 98 amino acid region of the antifreeze polypeptide from Hemitripterus americans, and a 27.0% amino acid identity over a 63 amino acid region of the lectin from Megabalanus rosa.

The homologies suggest that the protein encoded by the Sal F 3R functions as a lectin or as a homologue of the human low affinity FcR for IgE. The latter homology is particularly important, as the human low affinity FcR for IgE is the same as CD23, a cell surface protein expressed on B lymphocytes which is of central importance in regulating B cell growth (Gordon, J. & Guy G.R. (1987), Immunol. Today, 8, 339).

Thus, the vaccinia virus protein encoded by Sal F 3R is thought to function as an agonist of the normal CD23 molecule, to restrict the growth and/or differentiation of B cells and thereby reduce the host immune response to infection by the virus. Therefore, deletion of this gene from the virus genome would enhance the host immune response to the virus. The consequence of this could be restriction of virus growth and hence attenuation. It is

also possible that the immune response to foreign proteins expressed by recombinant vaccinia viruses lacking this gene would be enhanced and the efficacy of such candidate vaccines improved. Expression of the authentic human CD23 protein in vaccinia recombinants that do or do not contain the vaccinia homologue of CD23 may also enhance the immunogenicity of recombinant vaccinia virus vaccines that express antigens from heterologous pathogens.

If the protein has alternative or additional functions as a lectin, it may play a role in the attachment of virus to the target cell. Thus, deletion of the functioning gene in this capacity results in virus attenuation since the ability of virus particles to infect cells would be diminished.

A mutant virus with the coding region of this gene interrupted and partially deleted has been constructed. A plasmid, pPROF was constructed by the ligation of the leftmost 3524 bp (SalI-EcoRI DNA fragment) of the vaccinia virus SalI F fragment into pUC13 that had been digested with EcoRI and SalI. This plasmid contains the entire coding region of Sal F 3R and was digested with NsiI, which cuts twice only, within the coding sequence (Figure 7). The digested DNA was treated with bacteriophage T4 DNA polymerase to create blunt ends, and the larger of the two fragments was purified by agarose gel electrophoresis. This fragment was ligated with a gel-purified DNA fragment containing the E.coli xanthine guanine phosphoribosyl transferase (Ecogpt) gene joined to the vaccinia virus 7.5K promoter sequence. The latter fragment was obtained by

digestion of plasmid pGpt07/14 (Boyle, D.B. and Coupar, B.E. H. Gene 65, 123-8 (1988)) with EcoR1, followed by treatment of the digested DNA with DNA polymerase (Klenow fragment) to create blunt ends and isolation of a 2.1kb DNA fragment. The ligated DNA was cloned into E.coli, and the resulting bacterial colonies screened for the presence of the desired plasmid with appropriate restriction enzymes. Through this procedure, (outlined in Figure 7) a plasmid, pSAD3G, was isolated in which 100bp of the Sal F 3R coding sequence was replaced by a functional copy of the Ecogpt gene under the control of the vaccinia virus 7.5K promoter.

Plasmid pSAD3G was transfected into CV-1 cells that were infected with wild type (WT) vaccinia virus and the virus progeny derived from these cells after 48 hours at 37°C were then plated on fresh CV-1 cells in the presence of mycophenolic acid (MPA), xanthine and hypoxanthine. These drugs permit the replication only of recombinant viruses which contain and express the Ecogpt gene (Boyle & Coupar 1988 supra; Falkner & Moss, J. Virol, <u>62</u>, 1849-54, 1988). After three rounds of plaque purification, the virus was amplified in larger cultures of CV-1 cells. Southern blot analysis of virus DNA confirmed that the Ecogpt gene was present at the predicted location in the virus genome, that no functional copy of the Sal F 3R gene remained and that no other virus genomic DNA re-arrangements had occurred (see Figure 8). Since a virus lacking the Sal F 3R gene is viable, these data established that the gene SalF ORF3 is non-essential for virus replication in vitro.

The likelihood that the inactivation of the Sal F 3R coding region would generate an attenuated virus, depends on whether or not the region is expressed during normal virus replication. To address this point, virus mRNA transcribed from this region during the early and late phase of infection was analysed by Northern blotting. The results are shown in Figure 9. A single-stranded radio-labelled DNA probe complementary only to the coding strand of Sal F 3R detected an early mRNA species of about 600 nucleotides. Late during infection, this mRNA was replaced by some RNA species of heterogeneous length which appear as a smear on the Northern blot. Due to the heterogeneous length of late vaccinia virus mRNA, it is possible that this represents either mRNA initiating from the Sal F 3R promoter or from further upstream. This data allows the conclusion that the gene Sal F 3R is certainly transcribed early and possibly also late during infection.

2. Sal F 9R

The nucleotide and amino acid sequence of this gene is shown in Figure 10. The coding region of the gene resides between nucleotides 4447 and 4821 from the left end of the SalI F fragment. The encoded protein has a predicted molecular weight of 13.6kD.

Figure 11 shows the amino acid sequence homology between the protein encoded by gene Sal F 9R and two superoxide dismutase (Cu-Zn) proteins from cow (i) (Steinman et al., (1974), J. Biol. Chem., 249, 7326-38) and man (ii) (Sherman, L. et al (1983), PNAS, 80, 5465-9). The

protein encoded by SalI F ORF9 has a 36.8% amino acid identity over a 57 amino acid region of bovine superoxide dismutase, and a 37.3% amino acid identity over a 59 amino acid region of human superoxide dismutase. Superoxide dismutase (SOD) is an enzyme that converts toxic oxidative free radicals (O_2^-) into oxygen and hydrogen peroxide. Following engulfment of microorganisms, phagocytic cells undergo an oxidative burst which produces O_2^- to cause destruction of the microorganism. The presence of SOD in the structure of a virus, or its expression shortly after infection, would provide a defence mechanism against this toxic radical O_2^- (by converting it into oxygen and hydrogen peroxide) and thereby enhance the survival and replication of vaccinia virus in host macrophages, a site in which poxviruses can survive and which facilitate the systemic spread of the virus (Fenner F. (1985), in "Virology". B.N. Field (Ed) pp. 661-684, Raven Press, New York). However, a thorough analysis of the amino acid structure of the predicted protein encoded by the gene Sal F 9R, shows that it lacks some critical amino acid residues that are involved in binding of the copper and zinc divalent cations in other SOD enzymes. On the basis of this it seems unlikely that the vaccinia virus SOD homologue has SOD-like enzyme activity, and novel virus-induced enzyme activity has not been detected in infected cells. However, the presence of this gene in the virus genome remains interesting, and suggests that other poxviruses might retain a functional SOD enzyme. For example, African Swine Fever Virus (ASFV) is a likely candidate to contain this

enzyme as it replicates efficiently in swine macrophages. Deletion of this gene from the viruses containing the identified gene sequence and which retain superoxide dismutase enzyme activity would result in virus attenuation due to a reduced ability of the virus to replicate within, and be disseminated by, macrophages.

3. Sal F 13R

Figure 4 shows the deduced sequence of the 227 amino acid ORF designated Sal F 13R. Figure 2 shows the position of SalI F ORF13 within the vaccinia virus genome.

Approximately 40 nucleotides upstream of the ATG codon at the beginning of the ORF and 20 nucleotides downstream of the termination codon there are sequences TTTTTGT and TTTTTAT, respectively, which represent termination signals for early transcription (Yuen, L. and Moss, B. (1987) Proc. Natl. Acad. Sci. USA., 84, 6417-6421). The next downstream T_5NT motif is located a further 540 nucleotides away within the promoter region of the DNA ligase gene and contains two overlapping termination signals within the sequence TTTTTTTAT. The location of these early transcriptional termination signals and the absence of the sequence TAAAT(G) (a late transcription initiation site (Rosel, J.L., Earl, P.L. and Moss, B. (1986) J. Virol., 60, 436-449; Hanggi, M., Bannwarth, W. and Stunnenberg, H.G. (1986) EMBO J., 5, 1071-1076)) at the 5' ends of Sal F 13R suggests that the gene may be transcribed early during infection.

The coding region of the Sal F 13R gene maps between nucleotides 6313 and 7113 from the left end of the SalI F fragment. The encoded protein has a predicted molecular weight of 26.1kD.

The deduced amino acid sequence of Sal F 13R was compared against protein database SWISSPROT and our own database of vaccinia virus proteins using programme FASTP (Lipman, D.J. and Pearson, W.R. (1985) Science, 227, 1435-1441). No strong matches were found against other vaccinia proteins but the deduced amino acid sequence of SalF 13R had a high FASTP homology score (371) against thymidylate kinase (TmpK) of <u>Saccharomyces cerevisiae</u> (Jong, A.Y.S., Kuo, C.L. and Campbell, J.L. (1984) J. Biol. Chem., 259, 11052-11059; Rothstein, R., Helms, C and Rosenberg, N. (1987) Mol. Cell Biol. 7, 1198-1207).

Figure 12 shows the amino acid sequence homology between the protein encoded by gene Sal F 13R and thymidylate kinase (TmpK) from yeast (Jong et al, (1984), J. Biol. Chem. <u>259</u>, 11052-9).

The two proteins share 42% amino acid identity over a 200 amino acid region and there are many additional conservative changes. The aligned amino acid sequences are very similar in length, (yeast 216 amino acids versus vaccinia virus 204 amino acids), and are almost colinear. The computer predicted extra amino acid residues at the amino terminus of Sal F 13R which are upstream of the 5' end of early mRNA, have no homology with yeast TmpK. This is consistent with these amino acids not being part of the vaccinia TmpK enzyme. An alignment of the two amino acid

sequences is shown in Figure 13 with identical amino acids boxed.

Amino acids residues 11-18 of the putative vaccinia TmpK enzyme fit the consensus motif for ATP binding proteins GxxGxGKS/T (Otsuka, M. and Kit, S. (1984) Virology, 135, 316-330) except for the second glycine, where there is lysine. An alignment of this region with the presumed ATP binding sites of yeast TmpK (Jong, A.Y.S., Kuo, C.L. and Campbell, J.L. (1984) J. Biol. Chem., 259, 11052-11059; Rothstein, R., Helms, C. and Rosenberg, N. (1987) Mol. Cell Biol. 7, 1198-1207) HSV thymidine kinase (TK) TmpK (Otsuka, M. and Kitt, S. (1984) Virology, 135, 316-330; McKnight, S.L. (1980) Nucleic Acids Res. 8, 5949-5964; Wagner, M.J. Sharp, J.A. and Summers, W.C. (1981) Proc. Natl. Acad. Sci. USA., 78, 1441-1445; Gompels, U. and Minson, A.C. (1986) Virology, 153, 23-247; Darby, G., Larder, B.A. and Inglis, M.M. (1986) J. Gen. Virol., 67, 753-758; Kit, S., Kit. M., Qavi, H., Trkula, D and Otsuka, H. (1983) Biochem. Biophys. Acta, 741, 158-170; Swain, M.A. and Galloway, D.A. (1983) J. Virol, 46, 1045-1050) vaccinia TK (Weir, J.P. and Moss. B. (1983) J. Virol. 46, 530-537) and human TK (Bradshaw, H.D. and Deininger, P.L. (1984) Mol. Cell Biol., 4, 2316-2320) is shown in Figure 14A. In all these sequences the glycine residues at positions five and ten, lysine at position eleven and threonine at position thirteen are invariant. Only herpes simplex virus (HSV) TK/TmpK contains the second glycine of the ATP binding site consensus (above). The alignment of this region also shows that the highly homologous yeast and

vaccinia TmpK sequences and the more divergent HSV TK/TmpK, differ from TK sequences, of which vaccinia and man are representative examples, in several respects. First, immediately preceding the first glycine all with TmpK enzymes contain an acidic residue while TKs contain a hydrophobic residue. Second, at positions six to eight all poxvirus (Weir, J.P. and Moss. B. (1983) J. Virol, 46, 530-537; Boyle, D.B., Coupar, B.E.H., Gibbs, A.J., Seigman, L.J. and Both. G.W. (1987) Virology, 156, 335-367; Esposito, J.J. and Knight, J.C. (1984) Virology, 135, 561-567; Upton. C. and McFadden, G. (1986) J. Virol., 60, 920-927) and cellular (Bradshaw, H.D. and Deininger, P.L. (1984) Mol., Cell Biol., 4, 2316-2320; Lin. P.F. Lieberman, H.B., Yeh, D.B., Xu T., Zhao. S.Y. and Ruddle. F.H. (1985) Mol. Cell Biol,. 5, 3149-3156; Kwoh, T.J. and Engler, J.A. (1984) Nucleic Acids Res., 12, 3959-3971) and cellular TK enzymes contain PMF residues while yeast and vaccinia TmpK sequences contain LDK/R. Here the HSV enzyme fits neither pattern and this may reflect its broader substrate specificity. Third, as position fourteen poxvirus and cellular TKs contain glutamic acid while vaccinia and yeast TmpK contain glutamine and HSV has threonine.

Outside the ATP binding site there is no detectable homology between the vaccinia TmpK and TK sequences. However, homology exists between vaccinia TmpK and HSV TmpK/TK at a second nucleotide/nucleoside binding region. The alignment of the sequences from yeast TmpK, vaccinia TmpK and HSV TK/TmpK in this region is shown in Figure 14B. Although the yeast and vaccinia enzymes are clearly more

homologous, a TLI triplet is conserved between vaccinia and HSV (positions three to five).

The gene encoding TK has been mapped and sequenced. It is a nonessential gene for in vitro replication and has been widely used as a site for insertion of foreign DNA into recombinant vaccinia viruses (Mackett, M. and Smith. G.L. (1986) J. Gen. Virol. 67, 2067-2082). It is also a determinant of virus pathogenicity for both vaccinia (Buller, R.M.L., Smith. G.L., Cremer, K., Notkins, A.L. and Moss. B. (1985) Nature, 317, 813-815) and HSV (Field, H.J. and Wildy, P. (1978) J. Hyg. Camb. 81, 267-277; Kit S., Qavi, H., Dubbs, D.R. and Otsuka, H. (1983) J. Med. Virol., 12, 25-36). Deletion of the TK gene results in virus attenuation (Buller et al (1985) Nature, 317, 813-5).

The enzyme TmpK converts thymidine monophosphate (thymidylate or dTMP) into thymidine diphosphate (dTDP) within the biochemical pathway illustrated in Figure 15. Vaccinia virus encodes a separate enzyme, thymidine kinase (TK) that acts to convert thymidine into thymidine monophosphate in the first part of this pathway. Given that TK and TmpK perform sequential steps in the same biochemical pathway the present applicants have realised that very probably the vaccinia TmpK gene is also not essential for virus replication and that its deletion would also cause virus attenuation. This gene would therefore provide an additional site for insertion of foreign DNA into vaccinia virus and be a target for effecting virus attenuation.

Two vaccinia virus mutants have been constructed in

which the Sal F 13R gene has been inactivated. The Ecogpt gene joined to the vaccinia virus promoter p7.5K was inserted into a region of the TmpK gene predicted to be involved in nucleoside/nucleotide binding and, therefore, likely to be essential for enzyme activity (Smith et al. Nucleic Acids Res., 17, 7581, (1989)). The strategy followed that described above for the Sal F 13R gene (Figure 16). A plasmid, pACV1, was constructed by the ligation of a 2392 bp DraI DNA fragment, derived by DraI digestion of the SalI F fragment, into SmaI cut pUC13. pACV1 contains the entire Sal F 13R coding sequence and was digested with restriction enzyme MluI which cuts pACV1 only once and within the coding region of TmpK. The Ecogpt gene joined to the vaccinia virus promoter p7.5K was isolated as an EcoRI fragment (as above), made blunt-ended by treatment with DNA polymerase (Klenow fragment), and ligated with pACV1 that had been digested with MluI. The resultant plasmid, pACV2, contained the TmpK gene interrupted by Ecogpt. The procedure is outlined in Figure 16. This plasmid was used to transfect CV-1 cells infected with either WT vaccinia virus or a TK^- recombinant virus which expresses the hepatitis B virus surface antigen gene (Smith et al., Nature 302, 490-5, 1983). Recombinant viruses expressing the Ecogpt gene were selected by plaque assay in the presence of MPA and stocks grown. The virus derived from WT virus was called vAC1 and the virus derived from vHBs4 was called vACHB. Their genomic DNAs were analysed by Southern blotting (Figure 17). These data showed that both viruses contain the Ecogpt gene integrated

at the predicted location, that no other genomic alterations had occurred and established that the product of SalF 13R is non-essential for virus replication in vitro.

Transcriptional mapping by Northern blotting and S1 nuclease protection demonstrated that the SalF 13R gene is transcribed early but not late during infection. An early mRNA of approximately 850 nucleotides was detected with a probe specific for the coding strand of SalF 13R. This size corresponds to the size of the mRNA predicted if transcription initiates just upstream of the ORF and terminated 50 nucleotides downstream of the first downstream early transcription termination signal. S1 nuclease mapping precisely located the 5' end of the early mRNA to just upstream of the second inframe ATG codon. This is roughly 65 nucleotides downstream of the first ATG codon and the protein is therefore 23 amino acids shorter than that previously predicted. (Smith et al., Nucl. Acids Res. 17, 7581-90).

Assays for TmpK activity in vaccinia virus-infected cells have been performed and enzyme activity has been detected. The assays consist of incubating extracts of mock or virus-infected cells with tritiated thymidylate, resolving the reaction products by thin layer chromatography (TLC) (to separate TMP, TDT and TTP) and counting the areas of the tritium in TLC corresponding to these compounds (Jong et al., J.B.C. 259, 11052-9 (1984).

However, because TmpK is an essential cellular enzyme the applicants have not yet demonstrated a difference

between the endogenous activity in uninfected cells and that present in vaccinia virus-infected cells. To overcome this difficulty the applicants have reconstructed the gene by polymerase chain reaction (PCR) using synthetic oligonucleotides, re-sequenced and cloned into plasmid pEMBLyex4 (Dente et al, Nuc. Acids Res. 11, 1645-55, 1983) designed for expression of genes in *Saccharomyces cerevisiae*. The plasmid is used to complement a yeast mutant, CDC8, that is deficient in TmpK activity (Jong et al 1984). Complementation of this yeast strain directly shows that the Sal F 13R gene encodes TmpK enzyme activity, and since the parent yeast strain has no endogenous TmpK activity, it is straightforward to demonstrate enzyme activity *in vitro* using extracts of these yeast cells.

4. B5R

The nucleotide and amino acid sequence of gene B5R are shown in Figure 18. The encoded protein has a predicted molecular weight of 35.1 kD and its coding region maps between nucleotides 6654 and 7604 from the left end of the SalI G fragment. The protein contains hydrophobic amino acid sequences near the amino- and carboxy-termini, indicating that the protein associates with cell membranes of the infected cell or virus particle. There are also three potential sites for N-linked glycosylation indicating the mature product is a glycoprotein.

Comparisons of the amino acid sequence with the SWISSPROT protein database established significant homologies with several proteins that belong to the

superfamily of complement control proteins and blood coagulation factors. The alignments of Figure 19 show the amino acid sequence homology between the protein encoded by gene B5R (SalI G ORF10) and coagulation factor XIII B chain, complement factor H precursor, complement C2 precursor, and complement C4B-binding protein precursor. The protein encoded by B5R has a 27.2% amino acid identity with a 246 amino acid region of coagulation factor XIII B chain, a 27.2% amino acid identity with a 125 amino acid region of complement factor H precursor, a 26.4% amino acid identity with a 178 amino acid region of complement C2 precursor, and a 24.6% amino acid identity with a 175 amino acid region of complement C4B-binding protein precursor. Within the proteins of this superfamily, there are repeated domains of roughly 60 amino acids. The vaccinia protein encoded by gene B5R possesses four such domains.

Vaccinia virus contains a gene encoding another protein, H3C 28K (Kotwal, G. & Moss, B. (1988), Nature, 335, 176) which shows homology with this superfamily of complement and blood coagulation proteins and which is non-essential for virus replication. The protein encoded by gene B5R is related to, but distinct from, this protein, with a 29% amino acid homology. The H3C 28K protein is more closely related to the complement C4B-binding protein than the protein encoded by gene B5R. Conversely, the protein encoded by gene B5R is more closely related to coagulation factor XIII than the H3C 28K protein is. Another significant difference between the proteins SalI G ORF10 and H3C 28K is illustrated by the hydrophobicity

profiles shown in Figure 20. The presence of an extra hydrophobic domain near the carboxy-terminus of the protein encoded by B5R, and which is not shown H3C 28K, indicates that the former would remain cell associated whilst the latter is known to be secreted (Kotwal, G. & Moss, B. (1988), Nature 335, 176.

The homologies given above, indicate that the protein encoded by B5R is likely to interfere with the normal processes of complement activation (the H3C 28K protein is also known to do this) or blood coagulation. Interference in complement-mediated cell lysis would enhance the virus survival. Similarly, the prevention of blood clotting around the site of infection would prevent containment of the infection and enhance virus dissemination.

Attempts to construct a virus deletion mutant by insertional inactivation with Ecogpt have proved unsuccessful and it seems likely, but not proven, that this gene is essential for virus replication.

Where a gene is essential for virus replication, the virus may still be attenuated by altering the gene product. Thus, since the encoded protein binds complement factors, the region of the protein specific for the binding can be altered whilst keeping the protein functional for virus replication.

Transcriptional analysis of the B5R gene by Northern blotting (Figure 21) showed the presence of an early mRNA of 1850 nucleotides. This size corresponds to the size of the mRNA predicted if transcription initiates just upstream of B5R and terminates 50 nucleotides downstream of the

first downstream early transcription termination signal. There are also late RNAs of heterogeneous length from this region. S1 nuclease analysis has shown that the B5R promoter is expressed both early and late during infection. Unlike the constitutively active 7.5K promoter which also has early and late transcriptional initiation sites, the B5R promoter has the early RNA start site upstream of the late start site. The late start site maps to within a conserved motif TAAAT.

The protein has been expressed as a fusion protein with ß-galactosidase in E.coli and is currently being expressed in the authentic form in CHO cells driven by the human cytomegalovirus immediate early promoter-enhancer.

If the vaccinia protein functions as an anti-coagulation factor it is possible that this protein, or a form from which the carboxy hydrophobic domain has been deleted, would be a useful reagent in preventing blood coagulation.

5. Sal F 15R

The nucleotide and amino acid sequence of this gene are shown in Figure 3. The coding region of the gene maps between nucleotides 7625 and 9280 from the left end of the SalI F fragment. The encoded protein has a predicted molecular weight of approximately 63.3KD.

Transcriptional mapping of the SalI ORF15 gene by Northern blotting, S1 nuclease protection and primer extension have demonstrated that the gene is expressed early during infection (Smith et al., Nuc. Acids Res. 17,

9051-62). Surprisingly, the 5' end of the mRNA maps to a sequence TAAATG that is a characteristic of late transcription start sites. (Rosel, J.L., Earl, P.L., Weir, J.P. & Moss, B. (1986) J. Virol. 60, 436-449; Hanggi, M., Bannwarth, W. & Stunnenberg, H.G. (1986) EMBO J. 5, 1071-1076). The 5' end of the mRNA determined by primer extension maps 5 nucleotides upstream of the 5' end determined by S1 nuclease protection. It is possible that there are 5' oligo-adenylate residues on this early mRNA, which hitherto have been considered solely as a characteristic of late mRNA's.

Comparison of the amino acid sequence of Sal F 15R with our database of vaccinia virus proteins using programme FASTP (Lipman, D.J. & Pearson, W.R. (1985) Science 227, 1435-1441) found no strong matches. However, a search of the protein database SWISSPROT revealed extensive homology to DNA ligase of Saccharomyces cerevisiae (Barker, D.G., White, J.H.M. & Johnson, L.H. (1985) Nucleic Acids Res. 13, 8323-8337). An optimised FASTP score of 527 was obtained (KTUP of 1) and the two proteins had 30% amino acid identity over a 412 amino acid region. A similar degree of homology exists between SalI F ORF15 and Saccharomyces pombe DNA ligase (Barker, D.G. White, J.H.M., Johnston, L.H. (1987) Eur. J. Biochem. 162, 659-667) although fission yeast S.pombe and the budding yeast S.cerevisiae are evolutionarily divergent. Only weak homology was detected with bacteriophage T4 and T7 and E.coli DNA ligases. An alignment of the amino acid sequences of DNA ligases from yeasts and vaccinia virus is

shown in Figure 22. This alignment shows that the amino-terminal region of the vaccinia protein is divergent from both yeast sequences and there are regions which are absent in vaccinia but present in both yeasts. The latter point is reflected in the predicted sizes of the proteins, with vaccinia DNA ligase (63.3 kD) being considerably smaller than DNA ligases of S.pombe (86.2 kD) and S.cerevisiae (84.8 kD). The yeast DNA ligases are also least conserved in the amino terminal region. In contrast, in the carboxy-terminal region the three sequences are almost colinear and have extensive amino acid identity and conservative changes. The presumed catalytic lysine at the ATP binding site (marked with asterisk) is conserved in all these sequences as well as in T4 and T7 DNA ligases. In the E.coli enzyme, which uses NAD rather than ATP as cofactor, this site is less conserved. The most highly conserved region is very close to the carboxy terminus and is rich in basic amino acids. Over a 16 amino acid region the vaccinia protein shares identity with S.pombe at 15 positions with S.cerevisiae at 14 positions with a conservative isoleucine to valine change at one of the two divergent amino acids. This region is also well conserved in T4 with 6 identical residues and several conservative changes. The high conservation of this region suggests it plays some critical role in DNA ligase function, and its basic composition is consistent with an interaction with the DNA substrate.

The data below provide direct evidence that vaccinia virus encodes a DNA ligase and supports early data

(Sambrook, J. & Shatkin, A.J. (1969) J. Virol. 4, 719-726) showing a 13-fold increase in DNA ligase activity in the cytoplasm of vaccinia virus-infected cells. Spadari (Spadari, S. (1976) Nucleic Acids Res. 3, 2155-2167) concluded that the increase in DNA ligase activity was probably not virus-encoded since the enzyme had similar biochemical characteristics to cellular DNA ligase I, but may be attributable to enhanced leakage of the nuclear enzyme into the cytoplasm of virus infected cells.

The amino acid sequence of this vaccinia enzyme is the first reported primary structure of a 'mammalian' DNA ligase. It is also the only example of a eukaryotic virus encoding a DNA ligase, although other large DNA viruses which replicate in the cytoplasm, such as African Swine Fever Virus, probably encode this enzyme. Although much is known of mammalian DNA ligases, the genes encoding these enzymes have not been mapped.

Vaccinia virus contains two other enzymes with DNA strand sealing activity (topoisomerase and nicking-joining enzyme) and models for virus DNA replication have been proposed which do not require a conventional DNA ligase (Moyer, R.W. & Graves, R.L. (1981) Cell 27, 391-401; Baroudy, B.M., Venkatesan, S. & Moss, B. (1982) Cell 28, 315-324). In one model the linear double stranded DNA genome with covalently closed hairpin ends is nicked on one strand near one, or both, terminal hairpins to provide a 3' OH from which polymerisation may initiate. Elongation proceeds around the terminal hairpin, down the linear genome and around the opposite hairpin to produce

concatemeric DNA molecules by a strand displacement mechanism. This model does not require, but may use, lagging strand synthesis.

Figure 23 shows that extracts from vaccinia virus infected cells contain a novel radio-labelled polypeptide of molecular weight approximately 61 kD after incubation with α-(^{32}P) ATP. This activity is detectable in both crude and partially purified extracts, at early (lane 2) and late (lane 3) times post infection. The size estimated by SDS-PAGE is in good agreement with that predicted from the amino acid composition of Sal F 15R, 63 kD which would be consistent with a lack of extensive post-translational modification. The extent of incorporation of radioactivity is much greater than that in mock infected cells, in which only a faint band of approximately 46 kD is visible (lane 1). This polypeptide is also present at reduced intensity in extracts from vaccinia virus infected cells. The 130 kD mammalian DNA ligase I, highly purified from calf thymus, is shown in lane 5. Mock infected cells contain no polypeptide which co-migrates with the 61 kD band in infected cell extracts, suggesting the appearance of this protein is a consequence of infection with vaccinia virus.

The 61 kD polypeptide has the properties expected of a DNA ligase (Figure 24). A phosphocellulose column fraction derived from extracts of vaccinia virus infected cells late in infection was incubated with α-(^{32}P) ATP and then excess ATP was removed using Sephadex G-25. The excluded protein was incubated with either no addition (lane 4), DNA ligase substrate (lane 5) or sodium

pyrophosphate (lane 6). The presence of DNA substrate allows the ligase reaction to proceed to completion, with a disappearance of (^{32}P)-AMP from the enzyme. Conversely, high concentrations of pyrophosphate drive the equilibrium back towards free enzyme and ATP, again with a consequent discharge of radioactivity from the polypeptide (Figure 24). This result indicates that the 61 kD polypeptide is a DNA ligase with DNA strand joining activity.

An assay which measures ligation of 30 mer (^{32}P)-dT oligodeoxynucleotides annealed to poly dA was used to determine whether an increase in DNA ligase activity could be detected upon vaccinia virus infection. Activity is represented by the appearance of labelled products corresponding to two ligated molecules of the dT oligonucleotide (n=2), trimers of dT (n=3) and further higher oligomers. The assay of bacteriophage T4 DNA ligase provides a standard for this activity. An increase in DNA ligase activity above the basal level measurable in mock infected cells is observed after vaccinia virus infection (Figure 25). The ligase activity early in infection (lanes 5 and 8) is only slightly greater than the cellular activity in mock infected cells (lanes 6 and 9), but by late in infection (lanes 7 and 10) the activity is substantially higher. Extracts prepared in this low salt extraction buffer (100 mM NaCl), have an appreciable portion of the total DNA ligase activity located in the pellet fraction after centrifugation (lanes 8-10) compared to the supernatant (lanes 5-7) but an increase in the salt concentration to 1M shifts the majority of the total DNA

ligase activity into the soluble fraction (data not shown).

Approximately one third (183 amino acids) of the protein encoded by Sal F 15R was cloned into the bacterial expression vector pEX3 (Stanley and Luzio, EMBO J. 3, 1429 (1984) as described before (Kerr and Smith 1989)). A rabbit polyclonal antiserum (pEX LIG) was raised against the resulting ß-galactosidase/Sal F 15R fusion protein. The pEX LIG antiserum immune-precipitated two virus polypeptides from extracts of cells labelled with (^{35}S)-methionine. 1.5-4h post infection (Figure 26A, lane 2). The upper band, of molecular weight approximately 61 kD on SDS-PAGE, is more intense than the lower, of approximately 54 kD. No protein is recognised in mock infected cells (lane 1). The portion of the gene inserted into the pEX LIG construct does not include the regions of strongest amino acid sequence homology with yeast DNA ligases, therefore cross-reaction with mammalian ligases might not be expected. Both polypeptides are early virus gene products as treatment of the cells with cytosine arabinoside, an inhibitor of DNA replication, does not affect their expression (data not shown). De novo synthesis of both proteins can be detected early in infection by pulse labelling with (^{35}S) methionine 1.5-2.5h post infection followed by immune-precipitation, but only slightly reduced levels are observed as late as 7-8h p.i. (data not shown).

The larger of the two polypeptides detected by immune-precipitation with the pEX LIG antiserum co-migrates with the DNA ligase-AMP adduct from vaccinia virus infected

cells on SDS-PAGE (Figure 26). The marginal difference in size between the (^{32}P) and (^{35}S)-labelled polypeptides which may be detected on electrophoresis to achieve maximum resolution is possibly due to the addition of the AMP moiety in the (^{32}P)-labelled protein.

The DNA ligase would be likely to be an essential gene if it was involved in DNA replication, in which case it would not be possible to select recombinant virus containing a specific deletion of this gene. An alternative approach to prove that Sal F 15R gene product was responsible for the increase DNA ligase activity in vaccinia virus infected cells was therefore chosen. This made use of the pEX LIG antiserum, raised against Sal F 15R encoded protein, in immune-precipitation experiments against the radio-labelled DNA ligase-AMP adduct. The pEX LIG antiserum can efficiently precipitate the (^{32}P)-labelled DNA ligase protein in extracts from vaccinia virus infected cells (Figure 27, lane 5), whereas pre-immune serum from the same rabbit (lane 4), or a non-specific immune serum raised against an unrelated pEX fusion protein (lane 6), do not recognise the 61 kD polypeptide (Figure 28). Control experiments indicate that neither purified calf thymus (lane 1) nor bacteriophage T4 DNA ligase (lane 2) can be immune-precipitated by the pEX LIG antiserum (data not shown). The immune-precipitation of the novel DNA ligase-AMP adduct by the antiserum raised against Sal F 15R encoded protein clearly demonstrates that this vaccinia virus gene encodes the observed DNA ligase activity.

These data have been published in 1989 (Smith et al.,

Nucl. Acids Res. 17, 9051; and Kerr and Smith, Nucl. Acids Res. 17, 9039).

To assess the commercial potential of vaccinia virus DNA ligase, the applicants chose to over-express the gene in E.coli. To achieve this, the gene was precisely engineered by polymerase chain reaction (PCR) (Figure 28). An oligonucleotide representing the 5' end of the coding strand (including extra 5' nucleotides to form BamHI and NdeI sites) and an oligonucleotide complementary to the coding strand roughly 150 nucleotides downstream were used in a PCR reaction with the SalL F fragment cloned into a plasmid vector as template. The PCR fragment was digested with SalI and BclI and cloned into pSK16 that had been cut with SalL and BclI, to form pSK17. pSK16 contains the whole DNA ligase gene inserted into the SmaI site of pUC13 and was constructed by the isolation of a ClaI to MluI fragment from the pSK13. pSK13 contains the 3.3kb EcoRI to SmaI fragment of the SalI F fragment cloned into pUC13. The PCR fragment was sequenced to confirm no mutations had been introduced by PCR. Finally, the whole DNA ligase gene was excised from pSK17 with SalI and EcoRI and cloned into bacterial expression vector pGMT7 (Rosenberg et al., Gene, 56, 125-135, 1987), that had been digested with SalI and EcoRI, downstream of the T7 RNA polymerase promoter, to form pSK18. Introduction of pSK18 into E.coli strains bearing an inducible T7 RNA polymerase gene, resulted in high levels of DNA ligase expression in the presence of the specific inducer IPTG. Crude lysates of these induced bacteria contained a novel polypeptide of 61 kDa that bound

AMP (Figure 29). The applicants conclude that the vaccinia virus DNA ligase is active in E.coli and that the bacterial strain constructed potentially provides a large supply of this commercial important enzyme.

A deletion mutant of vaccinia virus lacking DNA ligase has been produced by the same procedure used for the Sal F 3R and Sal F 13R. Plasmid pSK13 was digested with NruI (which cuts just downstream of the ligase methionine initiation codon) and BglII (which cuts 997 bp further downstream). The overhanging ends were made blunt-ended with DNA polymerase (Klenow fragment) and the larger of the two fragments ligated with the Ecogpt gene linked to the vaccinia virus 7.5K promoter. The latter had been isolated as an EcoRI fragment and made blunt-ended with DNA polymerase (Klenow fragment). The resulting plasmid in which 1kb of the DNA ligase gene had been replaced with the Ecogpt gene was called pSK14 and was used to transfect WT vaccinia virus infected CV-1 cells. Ecogpt expressing viruses were isolated by growth in MPA and the DNA of several isolates analysed by Southern blot (Figure 30). These data show that some of the virus isolates have lost the internal 1 kb region of the DNA ligase gene but are still able to grow well in tissue culture. Consistent with this observation, assays of vaccinia virus DNA ligase in virus-infected cells (by the method described in Kerr & Smith, Nuc. Acids Res. 17, 9039, 1989) showed the absence of detectable DNA ligase in those viruses which had lost the DNA ligase DNA (Figure 31). These data indicate (surprisingly) that the enzyme is non-essential for virus

replication *in vitro* and that the DNA ligase gene is an additional site into which foreign DNA may be inserted into the virus genome. It is also probable that although the DNA ligase gene is non-essential for virus replication in tissue culture, the replication of DNA ligase-deficient viruses will be impaired *in vivo* and that such viruses will be attenuated.

Applications

In a recombinant vaccinia virus vaccine for use either in vaccination programmes or for use as an immunogen in the preparation of antibodies, the gene encoding the immunogen is isolated and introduced into the virus vector by conventional genetic engineering techniques, and the virus vector is transferred into the host, e.g. humans or animals by vaccination. Where the recombinant virus vaccine is being used for antibody production, antibodies to the immunogen are either extracted from the host antiserum (or unpurified antiserum may be used) using standard techniques well known in the art. Monoclonal antibodies may also be prepared from the cells of the immunised animals using standard techniques well known in the art.

The peptides encoded by the amino acid sequences encoded by the nucleotide sequences provided may be produced using conventional genetic engineering techniques.

The gene sequences identified also provide sites for the insertion of 'foreign' gene sequences into the vaccinia virus genome and may cause virus attenuation due to inactivation of the vaccinia genes.

CLAIMS

1. A vaccinia virus vector wherein:

a) part or all of one or more of the following nucleotide sequences is deleted from the viral genome; and/or

b) one or more of said nucleotide sequences is inactivated by mutation or the insertion of foreign DNA; and/or

c) one or more of said nucleotide sequences is changed to alter the function of a protein product encoded by said nucleotide sequence;

which nucleotide sequences are sequences designated herein as

i) Sal F 3R

ii) Sal F 9R

iii) Sal F 13R

iv) B5R

v) Sal F 15R.

2. A vaccinia virus according to claim 1 which comprises DNA sequences encoding one or more heterologous polypeptides.

3. A vaccinia virus according to claim 2 wherein the DNA sequences encoding one or more heterologous polypeptides are inserted into one or more ligation sites created by deleting part or all of said one or more nucleotide

sequences.

4. A vaccinia virus according to claim 1 which has enhanced immunogenicity.

5. A vaccinia virus according to claim 4 wherein

a) part or all of one or more vaccinia nucleotide sequences causing immunosuppression are deleted from the viral genome; and/or

b) one or more of said vaccinia nucleotide sequences causing immunosuppression is inactivated by mutation or the insertion of foreign DNA; and/or

c) one or more of said vaccinia nucleotide sequences causing immunosuppression is changed to alter the function of a protein product encoded by said nucleotide sequence;

which nucleotide sequences are sequences designated herein as

i) Sal F 3R

ii) Sal F 9R

iii) Sal F 13R

iv) B5R

v) Sal F 15R.

6. A vaccinia virus according to claim 5 wherein the vaccinia nucleotide sequence is a sequence designated herein as Sal F 3R.

7. A vaccinia virus according to claim 4 which comprises

a DNA sequence encoding a heterologous polypeptide which potentiates the immune response.

8. A vaccinia virus according to claim 7 wherein the DNA sequence encodes CD23.

9. A vaccine which comprises a vaccinia virus vector according to claim 1.

10. A medicament which comprises a vaccinia virus vector according to claim 1.

11. A polypeptide encoded by part or all of any of said nucleotide sequences;

which sequences are designated herein as

i) Sal F 3R

ii) Sal F 9R

iii) Sal F 13R

iv) B5R

v) Sal F 15R

and alleles and derivatives of said polypeptide.

12. A polypeptide according to claim 11 which is encoded by a nucleotide sequence designated herein as Sal F 15R and which has activity as a DNA ligase.

13. A method of attenuating a vaccinia virus vector which comprises:

a) deleting part or all of one or more of the following nucleotide sequences from the viral genome; and/or

b) inactivating one or more of said nucleotide sequences by mutating said nucleotide sequences or by inserting foreign DNA; and/or

c) changing said one or more nucleotide sequences to alter the function of a protein product encoded by said nucleotide sequence;

which nucleotide sequences are sequences designated herein as:

i) Sal F 3R

ii) Sal F 9R

iii) Sal F 13R

iv) B5R

v) Sal F 15R.

14. A method which comprises using a vaccinia virus vector according to claim 1 to prepare a vaccine or a medicament.

15. A method of using a vaccinia virus vector according to claim 1 as an immunogen for the production of antisera, monoclonal antibodies, polyclonal antibodies or T cells with specificity for a heterologous peptide encoded by a DNA sequence inserted into the viral genome; which method comprises immunising an animal with said vaccinia virus vector.

16. Monoclonal antibodies, polyclonal antibodies, antisera and/or T cells obtained by use of the method of claim 15.

17. Diagnostic test kits comprising monoclonal antibodies, polyclonal antibodies, antisera and/or T cells according to claim 16.

18. A method of using part or all of the nucleotide sequence designated herein as Sal F 15R, or part or all of the amino acid sequence encoded by said nucleotide sequence, to identify polypeptides with activity as a DNA ligase.

Fig. 1.

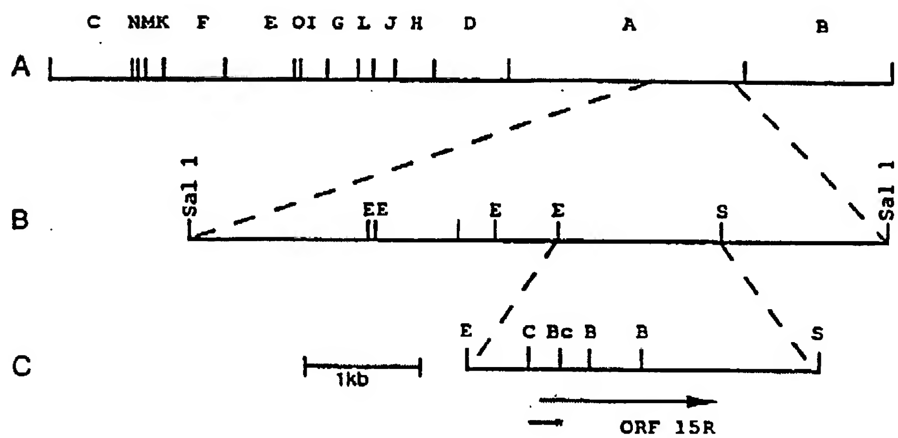

C NMK F E OI G L J H D A B
A
Sal 1
EE E E S
Sal 1
B
E C Bc B B S
C
1kb
ORF 15R

Fig. 2.

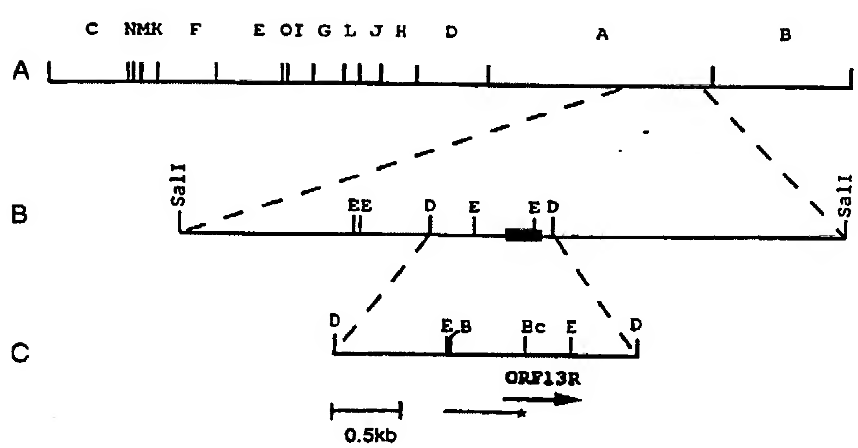

C NMK F E OI G L J H D A B
A
SalI
EE D E E D
SalI
B
D E B Bc E D
C
ORF13R
0.5kb

Fig.3.

```
CTCGGTTCCAGAATAGCCGATATAGATATTAAACAAGTGATACGTGAGAATATACTACAATTGTCTAATGATATCCGCGAACGATATTTGTGAAAAATATTAAAAAAAAATACTTTTTTT

      M  T  S  L  R  E  F  R  K  L  C  C  D  I  Y  H  A  S  G  Y  K  E  K  S  K  L  I  R  D  F  I  T  D  R  D  D  K  Y  L
ATTAAATGACGTCGCTTCGCGAATTTAGAAAATTATGCTGTGATATATATCACGCATCAGGATATAAAGAAAAATCTAAATTAATTAGAGACTTTATAACAGATAGGGATGATAAATATT

   I  I  K  L  L  L  P  G  L  D  D  R  I  Y  N  M  N  D  K  Q  I  I  K  L  Y  S  I  I  F  K  Q  S  Q  E  D  M  L  Q  D  L
TGATCATTAAGCTATTGCTTCCCGGATTAGACGATAGAATTTATAACATGAACGATAAACAAATTATAAAATTATATAGTATAATATTTAAACAATCTCAGGAAGATATGCTACAAGATT

   G  Y  G  Y  I  G  D  T  I  R  T  F  F  K  E  N  T  E  I  R  P  R  D  K  S  I  L  T  L  E  D  V  D  S  F  L  T  T  L  S
TAGGATACGGATATATAGGAGACACTATTAGGACTTTCTTCAAAGAGAACACAGAAATCCGTCCACGAGATAAAAGCATTTTAACTTTAGAAGACGTGGATAGTTTCTTAACTACGTTAT

   S  V  T  K  E  S  H  Q  I  K  L  L  T  D  I  A  S  V  C  T  C  N  D  L  K  C  V  V  M  L  I  D  K  D  L  K  I  K  A  G
CATCCGTAACTAAAGAATCGCATCAAATAAAATTATTGACTGATATCGCATCCGTTTGTACATGTAATGATTTAAAATGTGTAGTCATGCTTATTGATAAAGATCTAAAAATTAAAGCGG

   P  R  Y  V  L  N  A  I  S  P  N  A  Y  D  V  F  R  K  S  N  N  L  K  E  I  I  E  N  A  S  K  Q  N  L  D  S  I  S  I  S
GTCCTCGGTACGTACTTAACGCTATTAGTCCTAATGCCTATGATGTGTTTAGAAAATCTAATAACTTGAAAGAGATAATAGAAAATGCATCTAAACAAAATCTAGACTCTATATCTATTT

   V  M  T  P  I  N  P  M  L  A  E  S  C  D  S  V  N  K  A  F  K  K  F  P  S  G  M  F  A  E  V  K  Y  D  G  E  R  V  Q  V
CTGTTATGACTCCAATTAATCCCATGTTAGCGGAATCGTGTGATTCTGTCAATAAGGCGTTTAAAAAATTTCCATCAGGAATGTTTGCGGAAGTCAAATACGATGGTGAAAGAGTACAAG

   H  K  N  N  N  E  F  A  F  F  S  R  N  M  K  P  V  L  S  H  K  V  D  Y  L  K  E  Y  I  P  K  A  F  K  K  A  T  S  I  V
TTCATAAAAATAATAACGAGTTTGCCTTCTTTAGTAGAAACATGAAACCAGTACTCTCTCATAAAGTGGATTATCTCAAAGAATACATACCGAAAGCATTTAAAAAAGCTACGTCTATCG

   L  D  S  E  I  V  L  V  D  E  H  N  V  P  L  P  F  G  S  L  G  I  H  K  K  K  E  Y  K  N  S  N  M  C  L  F  V  F  D  C
TATTGGATTCTGAAATTGTTCTTGTAGACGAACATAATGTACCGCTCCCGTTTGGAAGTTTAGGAATACACAAAAAGAAAGAATATAAAAACTCTAACATGTGTTTGTTCGTGTTTGACT

   L  Y  F  D  G  F  D  M  T  D  I  P  L  Y  E  R  R  S  F  L  K  D  V  M  V  E  I  P  N  R  I  V  F  S  E  L  T  N  I  S
GTTTGTACTTTGATGGATTCGATATGACGGACATTCCATTGTACGAACGAAGATCTTTTCTCAAAGATGTTATGGTTGAAATACCCAATAGAATAGTATTCTCAGAGTTGACGAATATTA

   N  E  S  Q  L  T  D  V  L  D  D  A  L  T  R  K  L  E  G  L  V  L  K  D  I  N  G  V  Y  E  P  G  K  R  R  W  L  K  I  K
GTAACGAGTCTCAGTTAACTGACGTATTGGATGATGCACTAACGAGAAAATTAGAAGGATTGGTCTTAAAAGATATTAATGGAGTATACGAACCGGGAAAGAGAAGATGGTTAAAAATAA

   R  D  Y  L  N  E  G  S  M  A  D  S  A  D  L  V  V  L  G  A  Y  Y  G  K  G  A  K  G  G  I  M  A  V  F  L  M  G  C  Y  D
AGCGAGACTATTTGAACGAGGGTTCCATGGCAGATTCTGCCGATTTAGTAGTACTAGGTGCTTACTATGGTAAAGGAGCAAAGGGTGGTATCATGGCAGTCTTTCTAATGGGTTGTTACG

   D  E  S  G  K  W  K  T  V  T  K  C  S  G  H  D  D  N  T  L  R  V  L  Q  D  Q  L  T  M  V  K  I  N  K  D  P  K  K  I  P
ACGATGAATCCGGTAAATGGAAGACGGTTACCAAGTGTTCAGGACACGATGATAATACGTTAAGGGTTTTGCAAGACCAATTAACGATGGTTAAAATTAACAAGGATCCCAAAAAAATTC

   E  W  L  V  V  N  K  I  Y  I  P  D  F  V  V  E  D  P  K  Q  S  Q  I  W  E  I  S  G  A  E  F  T  S  S  K  S  H  T  A  N
CAGAGTGGTTGGTAGTTAATAAAATCTATATTCCCGATTTTGTAGTAGAGGATCCGAAACAATCTCAGATATGGGAAATTTCAGGAGCAGAGTTTACATCTTCCAAGTCCCATACCGCAA

   G  I  S  I  R  F  P  R  F  T  R  I  R  E  D  K  T  W  K  E  S  T  H  L  N  D  L  V  N  L  T  K  S
ATGGAATATCCATTAGATTTCCTAGATTTACTAGGATTAGAGAAGATAAAACGTGGAAAGAATCTACTCATCTAAACGATTTAGTAAACTTGACTAAATCTTAATAGTTACATACAAACT

GAAAATTAAAATAACACTATTTAGTTGGTGGTCGCCATGGATGGTGTTATTGTATACTGTCTAAACGCGTTAGTAAAACATGGCGAGGAAATAAATCATAT
```

3/29

Fig. 4.

```
                                                          M T Y I F I F S F I I R I N C K K Y V R   20
CTTAGATGGTTTAATATTTTTGTTACCCATGATCTATAAAGGTAGACCTAATCGTCTCGGATGACCTATATATTTATTTTCAGTTTTATTATACGCATAAATTGTAAAAAATATGTTAGG 120

F T K M S R G A L I V F E G L D K S G K T T Q C M N I M E S I P A N T I K Y L N   60
TTTACAAAAATGTCTCGTGGGGCATTAATCGTTTTTGAAGGATTGGACAAATCTGGAAAAACAACACAATGTATGAACATCATGGAATCTATACCGGCAAACACGATAAAATATCTTAAC 240
    *****
F P Q R S T V T G K M I D D Y L T R K K T Y N D H I V N L L F C A N R W E F A S  100
TTTCCTCAGAGATCCACTGTCACTGGAAAGATGATAGATGACTATCTAACTCGTAAAAAAACCTATAATGATCATATAGTTAATCTATTATTTTGTGCAAATAGATGGGAGTTTGCATCT 360

F I Q E Q L E Q G I T L I V D R Y A F S G V A Y A A A K G A S M T L S K S Y E S  140
TTTATACAAGAACAACTAGAACAGGGAATTACTTTAATAGTTGATAGATACGCGTTCTCTGGAGTAGCGTATGCCGCCGCTAAAGGCGCGTCAATGACTCTCAGTAAGAGTTATGAATCT 480

G L P K P D L V I F L E S G S K E I N R N V G E E I Y E D V T F Q Q K V L Q E Y  180
GGATTGCCTAAACCCGACTTAGTTATATTCTTGGAATCTGGTAGCAAAGAAATTAATAGAAACGTCGGCGAGGAAATTTATGAAGATGTTACATTCCAACAAAAGGTATTACAAGAATAT 600

K K M I E E G D I H W Q I I S S E F E E D V K K E L I K N I V I E A I H T V T G  220
AAAAAAATGATTGAAGAAGGAGATATTCATTGGCAAATTATTTCTTCTGAATTCGAGGAAGATGTAAAGAAGGAGTTGATTAAGAATATAGTTATAGAGGCTATACACACGGTTACTGGA 720

P V G Q L W M *                                                                                                          227
CCAGTGGGGCAACTGTGGATGTAATAGTGAAATTACATTTTTTATAAATAGATGTTAGTACAGTGTTATAAATGGATGAAG                                      800
```

Fig. 5.

```
TAGGGATAGGAGTAGGAATGCCACAAATGAAAAAAATACTTAAAATGTAATCTTAATCGAGTACACCACACGACA

 M  N  K  H  K  T  D  Y  A  G  Y  A  C  C  V  I  C  G  L  I  V  G  I  I  F
ATGAACAAACATAAGACAGATTATGCTGGTTATGCTTGCTGCGTAATATGCGGTCTAATTGTTGGAATTATTTTT

 T  A  T  L  L  K  V  V  E  R  K  L  V  H  T  P  S  I  D  K  T  I  K  D  A
ACAGCGACACTATTAAAAGTTGTAGAACGTAAATTAGTTCATACACCATCAATAGATAAAACGATAAAAGATGCA

 Y  I  R  E  D  C  P  T  D  W  I  S  Y  N  N  K  C  I  H  L  S  T  D  R  K
TATATTAGAGAAGATTGTCCTACTGACTGGATAAGCTATAATAATAAATGTATCCATTTATCTACTGATCGAAAA

 T  W  E  E  G  R  N  A  C  K  A  L  N  P  N  S  D  L  I  K  I  E  T  P  N
ACCTGGGAGGAAGGACGTAATGCATGCAAAGCTCTAAATCCAAATTCGGATCTAATTAAGATAGAGACTCCAAAC

 E  L  S  F  L  R  S  I  R  R  G  Y  W  V  G  E  S  E  I  L  N  Q  T  T  P
GAGTTAAGTTTTTTAAGAAGCATTAGACGCGGATATTGGGTAGGAGAATCCGAAATATTAAACCAGACAACCCCA

 Y  N  F  I  A  K  N  A  T  K  N  G  T  K  K  R  K  Y  I  C  S  T  T  N  T
TATAATTTTATAGCTAAGAATGCCACGAAGAATGGAACTAAAAAACGGAAATATATTTGTAGCACAACGAATACT

 P  K  L  H  S  C  Y  T  I
CCCAAACTGCATTCGTGTTACACTATATAACAATTACACTACATTTTTATCATACCACTACT
```

Fig.6.

(i)

```
HU FCR(IGE)    MEEGQYSEIEELPRRRCCRRGTQIVLLGLVTAALWAGLLTLLLLWHWDTTQSLKQLEERA
                       10        20        30        40        50        60

                                                                       10
SALF3R                                                         MNKHKTDYAGYA

HU FCR(IGE)    ARNVSQVSKNLESHHGDQMAQKSQSTQISQELEELRAEQQRLKSQDLELSWNLNGLQADL
                       70        80        90       100       110       120

                     20        30        40        50        60        70
SALF3R         CCVICGLIVGIIFTATLLKVVERKLVHTPSIDKTIKDAYIREDCPTDWISYNNKCIHLST
                            ...::. . :. : . ...  ...... ..X:..::....:: ....
HU FCR(IGE)    SSFKSQELNERNEASDLLERL-REEVTKLRMELQVSSGFVCNTCPEKWINFQRKCYYFGK
                      130       140        150       160       170

                     80        90       100         110       120       130
SALF3R         DRKTWEEGRNACKALNPNSDLIKIETPNELSFL--RSIRRGYWVGESEILNQTTPYNFIA
               . : : ..: :X...  ...:..:..:.: .::  .. . : :.:
HU FCR(IGE)    GTKQWVHARYACDDM--EGQLVSIHSPEEQDFLTKHASHTGSWIGLRNLDLKGEFIWVDG
            180       190         200       210       220       230

                    140       150
SALF3R         KNATKNGTKKRKYICSTTNTPKLHSCYTI

HU FCR(IGE)    SHVDYSNWAPGEPTSRSQGEDCVMMRGSGRWNDAFCDRKLGAWVCDRLATCTPPASEGSA
               240       250       260       270       280       290

HU FCR(IGE)    ESMGPDSRPDPDGRLPTPSAPLHS
               300       310       320
```

(ii)

```
                                        10        20        30        40
SALF3R                          MNKHKTDYAGYACCVICGLIVGIIFTATLLKVVERKLVHTPSI
                                                           :: .. . :.......
ANP            MQRQQADTETREDISTAGLSIIFIVCTISTTRMLTVSLLVCAMMALTQANDDKILKGTAT
                       10        20        30        40        50        60

                    50        60        70        80        90       100
SALF3R         DKTIKDAYIREDCPTDWISYNNKCIHLSTDRKTWEEGRNACKALNPNSDLIKIETPNELS
               . .  ..  . .X:..: . ...::. .:. .:X. . ..:. : ...: .:....:  :
ANP            EAGPVSQRAPPNCPAGWQPLGDRCIYYETTAMTWALAETNCMKL--GGHLASIHSQEEHS
                       70        80        90       100         110

                   110        120       130       140       150
SALF3R         FLRSIRRG-YWVGESEILNQTTPYNFIAKNATKNGTKKRKYICSTTNTPKLHSCYTI
               :..... :  :.:.:. :....
ANP            FIQTLNAGVVWIGGSACLQAGAWTWSDGTPMNFRSWCSTKPDDVLAACCMQMTAAADQCW
            120       130       140       150       160       170

ANP            DDLPCPASHKSVCAMTF
            180       190
```

Fig.6 (cont.)

(iii)

```
                  10        20        30        40        50        60
SALF3R    MNKHKTDYAGYACCVICGLIVGIIFTATLLKVVERKLVHTPSIDKTIKDAYIREDCPTDW
                                                                    X:
LEC                                                           TCPGNLDW

                  70        80        90       100        110
SALF3R    ISYNNKCIHLSTDRKTWEEGRNACKALNPNSDLIKIETPNELSFLR-SIRRG-YWVGESE
           .:...:   :: .  :.... :X....:.. : .:... : .:.. .....  :.: ..
LEC       QEYDGHCYWASTYQVRWNDAQLACQTVHPGAYLATIQSQLENAFISETVSNNRLWIGLND
            10        20        30        40        50        60

         120       130       140       150
SALF3R    ILNQTTPYNFIAKNATKNGTKKRKYICSTTNTPKLHSCYTI
          :
LEC       IDLEGHYVWSNGEATDFTYWSSNNPNNWENQDCGVVNYDTVTGQWDDDDCNKNKNFLC
          70        80        90        100       110       120

LEC       PIIGCPPCGI
         130
```

Fig. 7.

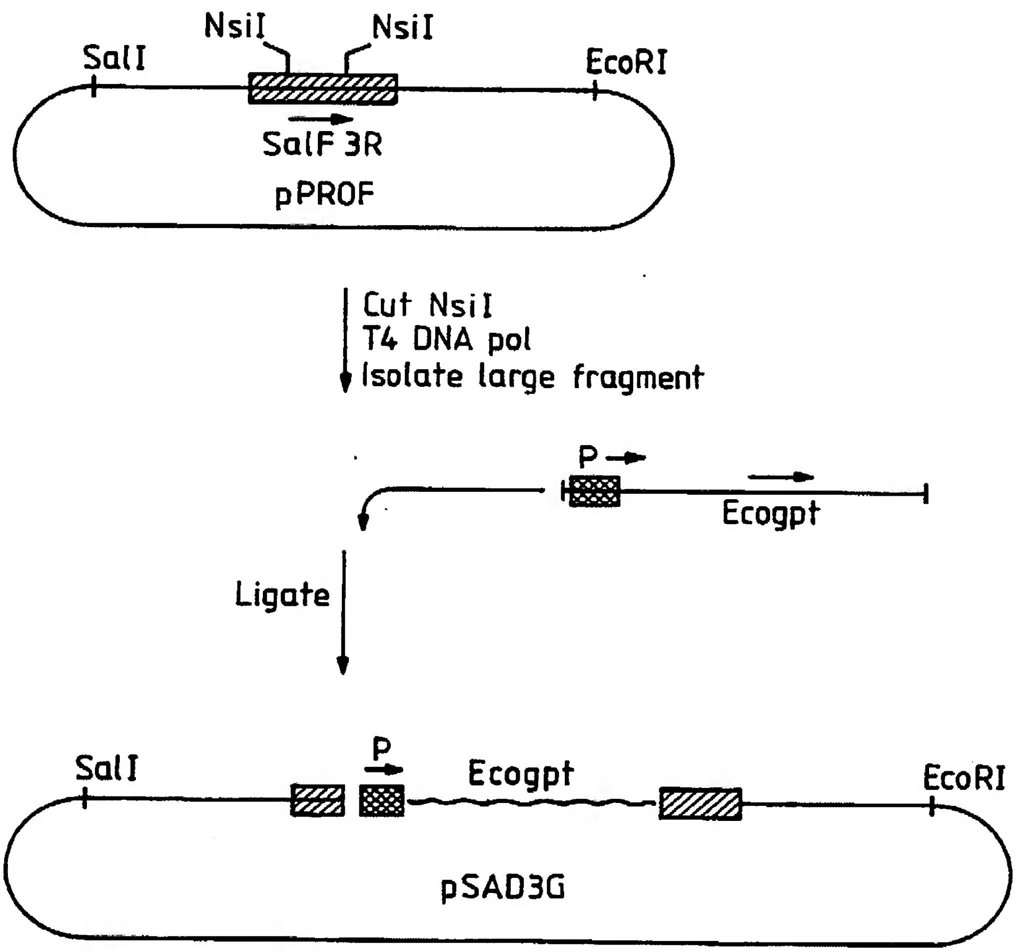

NsiI
NsiI
SalI
EcoRI
SalF 3R
pPROF
Cut NsiI
T4 DNA pol
Isolate large fragment
P
Ecogpt
Ligate
P
SalI
Ecogpt
EcoRI
pSAD3G

Fig.8.

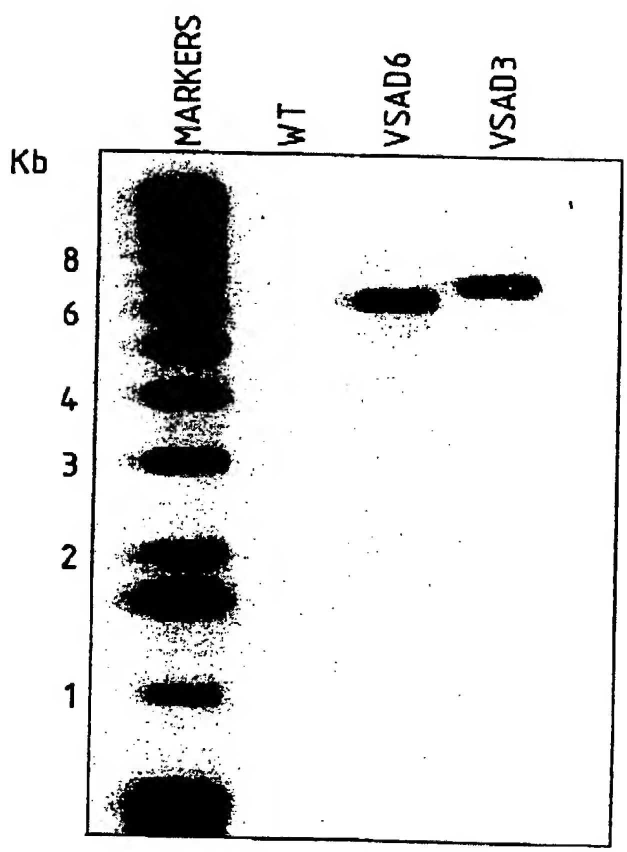


Fig. 9.

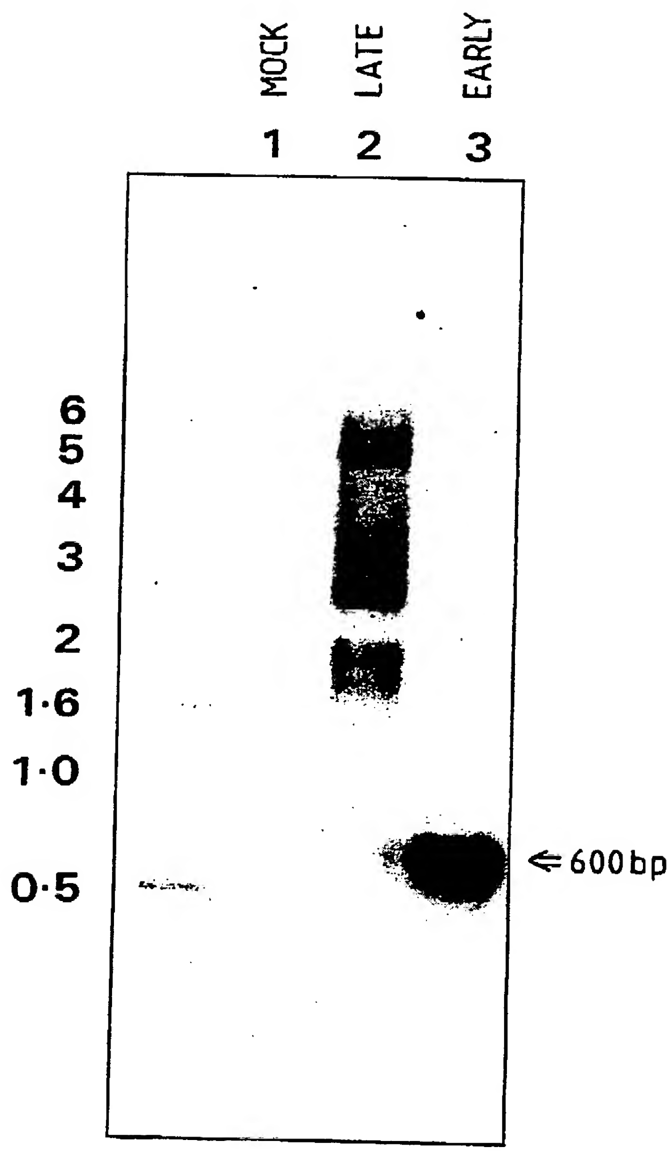


Fig. 10.

```
GCACCACCAGTAACCGCGTACACGGCCATTGCTGCCACTCATAATATCAGACTACTTATTCTATTTTACTAAATA

 M  A  V  C  I  I  D  H  D  N  I  R  G  V  I  Y  F  E  P  V  H  G  K  D  K
ATGGCTGTTTGTATAATAGACCACGATAATATCAGAGGAGTTATTTACTTTGAACCAGTCCATGGAAAAGATAAA

 V  L  G  S  V  I  G  L  K  S  G  T  Y  S  L  I  I  H  R  Y  G  D  I  S  Q
GTTTTAGGATCAGTTATTGGATTAAAATCCGGAACGTATAGTTTGATAATTCATCGTTACGGAGATATTAGTCAA

 G  C  D  S  I  G  S  P  E  I  F  I  G  N  I  F  V  N  R  Y  G  V  A  Y  V
GGATGTGATTCCATAGGCAGTCCAGAAATATTTATCGGTAACATCTTTGTAAACAGATATGGTGTAGCATATGTT

 Y  L  D  T  D  V  N  I  S  T  I  I  G  K  A  L  S  I  S  K  N  D  Q  R  L
TATTTAGATACAGATGTAAATATATCTACAATTATTGGAAAGGCGTTATCTATTTCAAAAAATGATCAGAGATTA

 A  C  G  V  I  G  I  S  Y  I  N  E  K  I  I  H  F  L  T  I  N  E  N  G  V
GCGTGTGGAGTTATTGGTATTTCTTACATAAATGAAAAGATAATACATTTTCTTACAATTAACGAGAATGGCGTT

TGATATATCAGTTAATGCGTCTAAAACAATAAATGCATTAGTTT
```

Fig. 12.

```
                     10        20        30        40        50        60
SALF13R     MTYIFIFSFIIRINCKKYVRFTKMSRGALIVFEGLDKSGKTTQCMNIMESIPANTIKYLN
                                   X.:: ::..::::..:::::: ::.  .    .  .  :  :.
sc TmpK                            MMGRGKLILIEGLDRTGKTTQC-NILYKKLQPNCKLLK
                                            10        20         30

                     70        80        90        100       110
SALF13R     FPQRSTVTGKMIDDYLTRKK-TYNDHIVNLLFCANRWEFASFIQEQLEQGITLIVDRYAF
            ::.::: .: .:..::: ..   .:. ..:::.:::::... :...: .: ....::::..
sc TmpK     FPERSTRIGGLINEYLTDDSFQLSDQAIHLLFSANRWEIVDKIKKDLLEGKNIVMDRYVY
             40        50        60        70        80        90

         120       130       140       150         160       170
SALF13R     SGVAYAAAKGA-SMTLSKSYES--GLPKPDLVIFL---ESGSKEINRNVGEEIYEDVTFQ
            :::::.::::. .:.:. . ..  :: ::::..::   . .... ... :.: ::.:.::
sc TmpK     SGVAYSAAKGTNGMDLDWCLQPDVGLLKPDLTLFLSTQDVDNNAEKSGFGDERYETVKFQ
            100       110       120       130       140       150

                 180       190       200       210       220
SALF13R     QKVLQEYKKMIEEGDIHWQIISSEFEEDVKKELIKNIVIEAIHTVTGPVGQLWM
            .:: :...:.... .:.       :  ..  .  ..:..
sc TmpK     EKVKQTFMKLLDK-EIRKGDESITIVDVTNKDIQEVEALIWQIVEPVLSTHIDHDKFSFF
            160       170       180       190       200       210
```

Fig. 11.

```
(i)
                   10        20        30        40        50
SALF9R    MAVCII-DHDNIRGVIYFEPVHGKDKVLGSVIGLKSGTYSLIIHRYGDISQGCDSIGS
          :::..  ...  ..:.:.X:.    ..    :  ::..::..:....  .:..::  .::X.:  :.
BOVSOD  ATKAVCVLKGDGPVQGTIHFEAKGDTVVVTGSITGLTEGDHGFHVHQFGDNTQGCTSAGP
                10        20        30        40        50        60

          60        70        80        90       100       110
SALF9R  PEIFIGNIFVNRYGVAYVYLDTDVNISTIIGKALSISKNDQRLACGVIGISYINEKIIHF

BOVSOD  HFNPLSKKHGGPKDEERHVGDLGNVTADKNGVAIVDIVDPLISLSGEYSIIGRTMVVHEK
                70        80        90       100       110       120

         120
SALF9R  LTINENGV

BOVSOD  PDDLGRGGNEESTKTGNAGSRLACGVIGIAK
               130       140       150

(ii)
                   10        20        30        40        50
SALF9R    MAVCII-DHDNIRGVIYFEP--VHGKDKVLGSVIGLKSGTYSLIIHRYGDISQGCDSI
          :::..  ...  ..:.: ::.     .:   X:  ::.  ::..:  ...  .:  .::  ..:X.:
HUMSOD  ATKAVCVLKGDGPVQGIINFEQKESNGPVKVWGSIKGLTEGLHGFHVHEFGDNTAGCTSA
                10        20        30        40        50        60

            60        70        80        90       100       110
SALF9R  GSPEIFIGNIFVNRYGVAYVYLDTDVNISTIIGKALSISKNDQRLACGVIGISYINEKII
        :.
HUMSOD  GPHFNPLSRKHGGPKDEERHVGDLGNVTADKDGVADVSIEDSVISLSGDHCIIGRTLVVH
                70        80        90       100       110       120

           120
SALF9R  HFLTINENGV

HUMSOD  EKADDLGKGGNEESTKTGNAGSRLACGVIGIAQ
               130       140       150
```

Fig. 13.

```
                                                                58
VV    MSRGALIVFEGLDKSGKTTQCMNIMESIPANTIKYINFPQRSTVTGKMIDDYLTRKK-T
SC    MMGRGKLILIEGLDRTGKTTQC-NILYKKLQPNCKLLKFPERSTRIGGLINEYLTDDSFQ
                                                                59
                                                               117
VV    YNDHIVNLLFCANRWEFASFIQEQIEQGITLIVDRYAFSGVAYAAAKGA-SMTLSKSYES
SC    LSDQAIHLLFSANRWEIVDKIKKDILEGKNIVMDRYVYSGVAYSAAKGTNCMDLDWCLQP
                                                               119
                                                               172
VV    --GIPKPDIVIFIESGSKEIN---RNVGEEIYEDVTFCQKVTCEYKKMIEEGDIHWQIIS
SC    DVGIILKPDITIFISTQDVDNNAEKSGFGDEPYETVKFQEKVKQTFMKLLDK-EIRKGDES
                                                               178
                              204
VV    SEFEEDVKKEIIKNIVIEAIHTVTGPVGQLWM
SC    ITI-VDVTNKDIQE-VEAIIWQIVEPVLSTHIDHDKFSFF
                                          216
```

Fig. 14.

A ATP BINDING SITE

VV TmpK	(7)	I	V	F	E	G	L	D	K	S	G	K	T	T	Q
SC TmpK	(8)	I	L	I	E	G	L	D	R	T	G	K	T	T	Q
HSV TK	(52)	V	Y	I	D	G	P	H	G	M	G	K	T	T	T
VV TK	(7)	Q	L	I	I	G	P	M	F	S	G	K	S	T	E
MAN TK	(22)	Q	V	I	L	G	P	M	F	S	G	K	S	T	E

B NUCLEOTIDE/NUCLEOSIDE BINDING SITE

VV TmpK	(86)	G	I	T	L	I	V	D	R	Y	A	F	-	S	G	V	A	Y	A	A	A	K	G
SC TmpK	(87)	G	K	N	I	V	M	D	R	Y	V	Y	-	S	G	V	A	Y	S	A	A	K	G
HSV TK	(156)	A	L	T	L	I	F	D	R	H	P	I	A	A	L	L	C	Y	P	A	A	R	Y

Fig. 15.

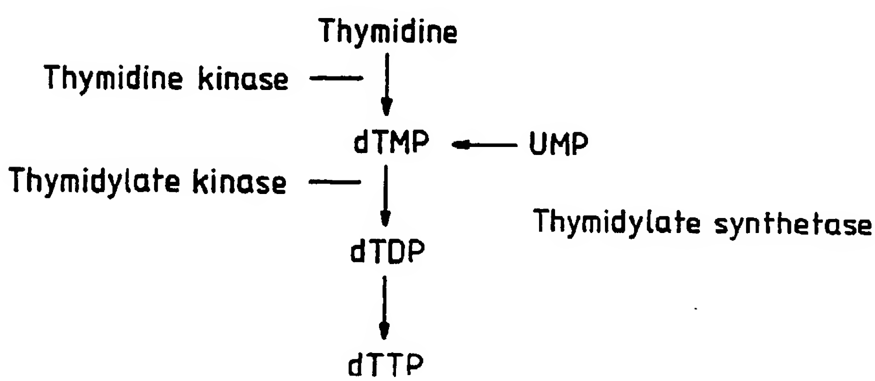


Fig. 16.

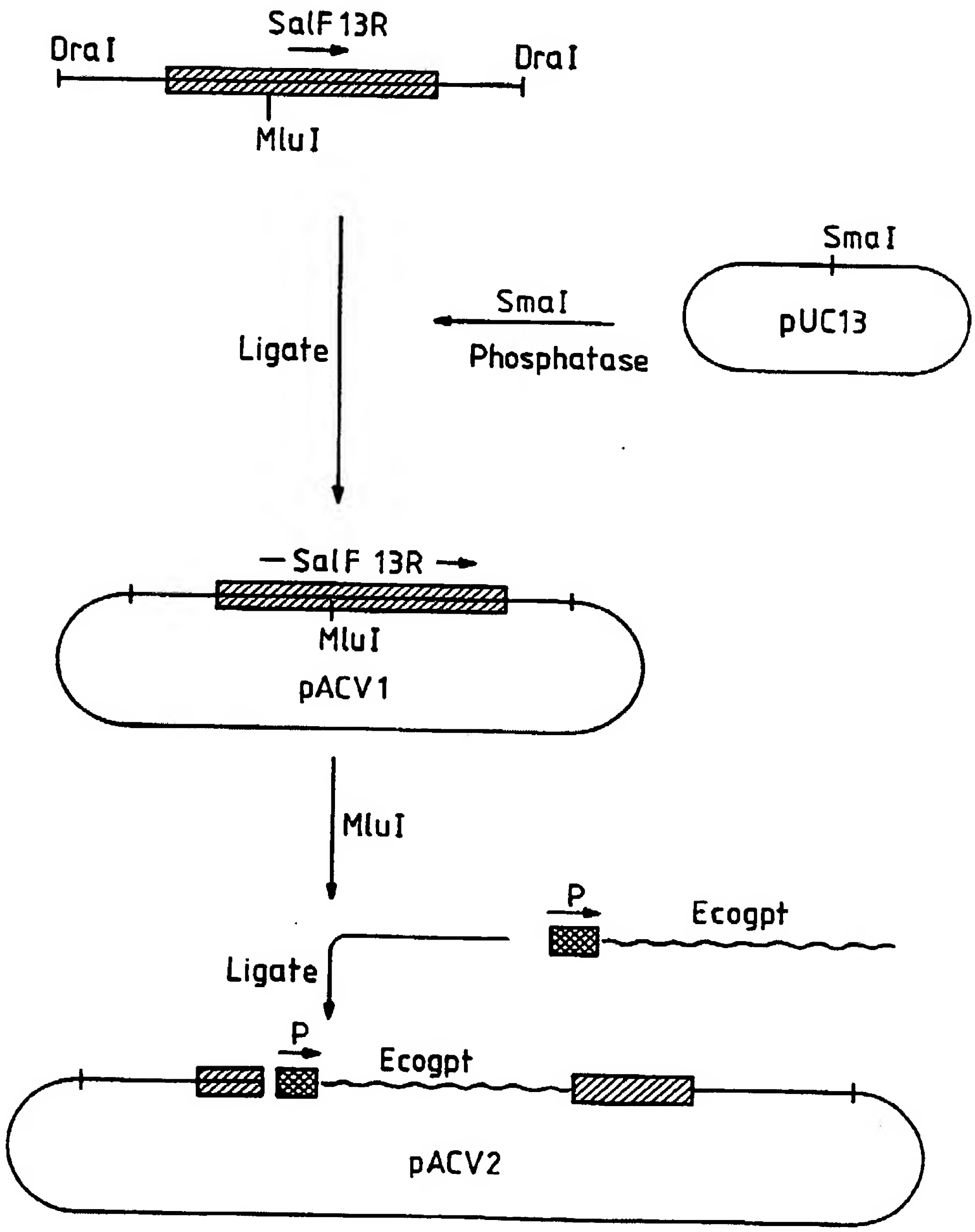

SalF 13R
DraI
DraI
MluI
SmaI
Ligate
SmaI
Phosphatase
pUC13
SalF 13R
MluI
pACV1
MluI
P
Ecogpt
Ligate
P
Ecogpt
pACV2

Fig. 17.

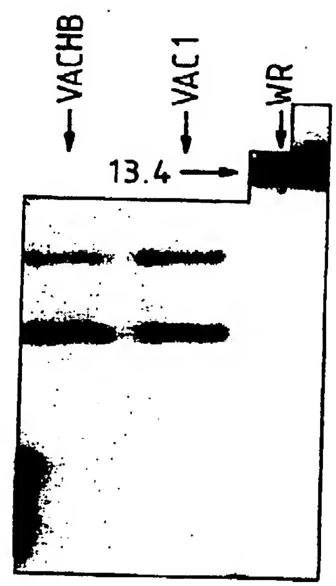


Fig.18.

GAACTATCCACATATCTAAAAATCTTATAAACACTATTAAAATATAAAATCTAAGTAGGATAAAATCACACTACATCATTGTTTCCTTTTAGTGCTCGACAGTGTATACTATTTTTAACA

M K T I S V V T L L C V L P A V V Y S T C T V P T M N N A K L T S T E T
CTCATAAATAAAAATGAAAACGATTTCCGTTGTTACGTTGTTATGCGTACTACCTGCTGTTGTTTATTCAACATGTACTGTACCCACTATGAATAACGCTAAATTAACGTCTACCGAAAC

S F N D K Q K V T F T C D Q G Y H S S D P N A V C E T D K W K Y E N P C K K M C
ATCGTTTAATGATAAACAGAAAGTTACGTTTACATGTGATCAGGGATATCATTCTTCGGATCCAAATGCTGTCTGCGAAACAGATAAATGGAAATACGAAAATCCATGCAAAAAAATGTG

T V S D Y I S E L Y N K P L Y E V N S T M T L S C N G E T K Y F R C E E K N G N
CACAGTTTCTGATTACATCTCTGAATTATATAATAAACCGCTATACGAAGTGAATTCCACCATGACACTAAGTTGCAACGGCGAAACAAAATATTTTCGTTGCGAAGAAAAAAATGGAAA

T S W N D T V T C P N A E C Q P L Q L E H G S C Q P V K E K Y S F G E Y M T I N
TACTTCTTGGAATGATACTGTTACGTGTCCTAATGCGGAATGTCAACCTCTTCAATTAGAACACGGATCGTGTCAACCAGTTAAAGAAAAATACTCATTTGGGGAATATATGACTATCAA

C D V G Y E V I G A S Y I S C T A N S W N V I P S C Q Q K C D M P S L S N G L I
CTGTGATGTTGGATATGAGGTTATTGGTGCTTCGTACATAAGTTGTACAGCTAATTCTTGGAATGTTATTCCATCATGTCAACAAAAATGTGATATGCCGTCTCTATCTAATGGATTAAT

S G S T F S I G G V I H L S C K S G F T L T G S P S S T C I D G K W N P V L P I
TTCCGGATCTACATTTTCTATCGGTGGCGTTATACATCTTAGTTGTAAAAGTGGTTTTACACTAACGGGGTCTCCATCATCCACATGTATCGACGGTAAATGGAATCCCGTACTCCCAAT

C V R T N E E F D P V D D G P D D E T D L S K L S K D V V Q Y E Q E I E S L E A
ATGTGTACGAACTAACGAAGAATTTGATCCAGTGGATGATGGTCCCGACGATGAGACAGATTTGAGCAAACTCTCGAAAGACGTTGTACAATATGAACAAGAAATAGAATCGTTAGAAGC

T Y H I I I V A L T I M G V I F L I S V I V L V C S C D K N N D Q Y K F H K L L
AACTTATCATATAATCATAGTGGCGTTAACAATTATGGGCGTCATATTTTTAATCTCCGTTATAGTATTAGTTTGTTCCTGTGACAAAAATAATGACCAATATAAGTTCCATAAATTGCT

P
ACCGTAAATATAAATCCGTTAAAATAATGAATAATTAATAACGAACAAGTATCAAAAGATTAAAGAA

Fig. 19. (i)

```
               10        20        30        40        50        60
B5R    MKTISVVTLLCVLPAVVYSTCTVPTMNNAKLTSTETSFNDKQKVTFTCDQGYHSSDPNAV
                           :: .:.. :.. ..:...:. :.:: ..:. ::.... . .
F13B   VQCLSDGWSSQPTCRKEHETCLAPELYNGNYSTTQKTFKVKDKVQYECATGYYTAGGKKT
             140       150       160       170       180       190

                     70        80        90           100       110
B5R    ----CETDKWKYENPCKKM-CTVSDYISELYNKPL---YEVNSTMTLSCNGETKYFRCEE
           : :   :. .   :.:. :.       :.. : .:.     :: .... . :. :. :.. ..
F13B   EEVECLTYGWSLTPKCTKLKCSSLRLIENGYFHPVKQTYEEGDVVQFFCH-ENYYLSGSD
             200       210       220       230       240       250

              120         130         140       150       160
B5R    KNG--NTSW-NDTVTC--PNAECQPLQLE-HGSCQPVKEKYSFGEYMTINCDVGYEVIGA
             : .:    .. .:    ... :.: .:   ... :.  ...X. ::  . :.:....:. :.
F13B   LIQCYNFGWYPESPVCEGRRNRCPPPLPINSKIQTHSTTYRHGEIVHIECELNFEIHGS
              260       270       280       290       300       310

        170       180            190        200       210
B5R    SYISCTANSWNVIPSC---QQK--CDMPSL---SNGLISGSTFSIGGVIHLSCKSGFTLT
       . :.:....:.  :.X    :.:  :. :..    ... . ....  :. .   .::::. :
F13B   AEIRCEDGKWTEPPKCIEGQEKVACEEPPFIENGAANLHSKIYYNGDKVTYACKSGYLLH
              320       330       340       350       360       370

      220       230       240       250       260       270
B5R   GSPSSTCIDGKWNPVLPICVRTNEEFDPVDDGPDDETDLSKLSKDVVQYEQEIESLEATY
      :: .  ::  :::  .. : :: .:.
F13B  GSNEITCNRGKW-TLPPECVENNENCKHPPVVMNGAVADGILASYATGSSVEYRCNEYYL
              380       390       400       410       420

      280       290       300       310
B5R   HIIIVALTIMGVIFLISVIVLVCSCDKNNDQYKFHKLLP

F13B  LRGSKISRCEQGKWSSPPVCLEPCTVNVDYMNRNNIEMK
    430       440       450       460
```

(ii)

```
              10        20        30        40        50        60
B5R   MKTISVVTLLCVLPAVVYSTCTVPTMNNAKLTSTETSFNDKQKVTFTCDQGYHSSDPNAV

CFAH  SCPPPPQIPNTQVIETTVKYLDGEKLSVLCQDNYLTQDSEEMVCKDGRWQSLPRCIEKIP
       810       820       830       840       850       860

              70        80        90       100       110       120
B5R   CETDKWKYENPCKKMCTVSDYISELYNKPLYEVNSTMTLSCNGETKYFRCEEKNGNTSWN
                                                                 :.
CFAH  CSQPPTIEHGSINLPRSSEERRDSIESSSHEHGTTFSYVCDDGFRIPEENRITCYMGKWS
       870       880       890       900       910       920

             130       140       150       160       170
B5R   DTVTCPNAEC-QPLQLEHGSCQPVKEKYSFGEYMTINCDVGYEVIGASYISCTANSWNVI
      ..   :  .   : .:   .   :.       X:.:  :: .: .:..:... :...: :....:.
CFAH  TPPRCVGLPCGPPPSIPLGTVSLELESYQHGEEVTYHCSTGFGIDGPAFIICEGGKWSDP
       930       940       950       960       970       980

    180        190        200       210       220       230
B5R   PSC-QQKCD-MPSLSNGLISG---STFSIGGVIHLSCKSGFTLTGSPSSTCIDGKWNPVL
      :.X  . .::  .:...:..:.:       .....:. .  ..:.: .  ..:: . ::....:
CFAH  PKCIKTDCDVLPTVKNAIIRGKSKKSYRTGEQVTFRCQSPYQMNGSDTVTCVNSRW-IGQ
       990      1000      1010      1020      1030      1040

          240       250       260       270       280       290
B5R   PICVRTNEEFDPVDDGPDDETDLSKLSKDVVQYEQEIESLEATYHIIIVALTIMGVIFLI
      :.:
CFAH  PVCKDNSCVDPPHVPNATIVTRTKNKYLHGDRVRYECNKPLELFGQVEVMCENGIWTEKP
       1050      1060      1070      1080      1090      1100

          300       310
```

Fig. 19 (cont.)

```
B5R    SVIVLVCSCDKNNDQYKFHKLLP

CFAH   KCRDSTGKCGPPPPIDNGDITSL
        1110         1120

(iii)
               50        60        70        80        90       100
B5R    QKVTFTCDQGYHSSDPNAVCETDKWKYENPCKKMCTVSDYISELYNKPLYEVNSTMTLSC

CO2    MGPLMVLFCLLFLYPGLADSAPSCPQNVNISGGTFTLSHGWAPGSLLTYSCPQGLYPSPA
               10        20        30        40        50        60

              110       120       130       140       150       160
B5R    NGETKYFRCEEKNGNTSWNDTVTCPNAEC-QPLQLEHGSCQPVKEKYSFGEYMTINCDVG
                 .. :.:.  ....:  . ^v.:. .:.: . :  ..:. :. ....:. :
CO2    SRLCKSSGQWQTPGATRSLSKAVCKPVRCPAPVSFENGIYTPRLGSYFVGGNVSFECEDG
               70        80        90       100       110       120

              170       180        190       200       210
B5R    YEVIGASYISCTANS-WNV-IPSCQQ---KCDMPSLSNGLI-SGSTFSIGGVIHLSCKSG
       . . :..   : .:X :.  .. :..   .:  :..: : . .:  :. :. .. .:.:.
CO2    FILRGSPVRQCRPNGMWDGETAVCDNGAGHCPNPGISLGAVRTGFRFGHGDKVRYRCSSN
              130       140       150       160       170       180

           220       230       240        250       260       270
B5R    FTLTGSPSSTCI-DGKWNPVLPICVRTNEEFD-PVDDGPDDETDLSKLSKDVVQYEQEIE
       ..::::....:   .: :. . ::: :    ..: : : .:. .:..:..  .. . ... :
CO2    LVLTGSSERECQGNGVWSGTEPIC-RQPYSYDFPEDVAPALGTSFSHMLGATNPTQKTKE
              190       200       210       220       230

            280       290       300       310
B5R    SLEATYHIIIVALTIMGVIFLISVIVLVCSCDKNNDQYKFHKLLP
       ::. . .:
CO2    SLGRKIQIQRSGHLNLYLLLDCSQSVSENDFLIFKESASLMVDRI
     240      250       260       270       280

(iv)
              10        20        30        40        50
B5R    MKTISVVTLLCVLPAVVYSTCTVPTMNNAKLTSTETSFNDKQKVTFTCDQ-GYHSSDP--
                              :. :..:.:..:   .:   ...  : :  :. .. . ..
C4BP   TVICQKNLRWTPYQGCEALCCPEPKLNNGEITQHRKSRPANHCVYFYGDEISFSCHETSR
        300       310       320       330       340       350

         60        70        80         90        100       110
B5R    -NAVCETD-KWKYENP-CKKMCTVSDYISE-LYNKP-LYE-VNSTMTLSCN-GETKYFRC
        .:.:..: .:.  .: : ..:. .  :..  :...  :.   ..... .:. : .    .
C4BP   FSAICQGDGTWSPRTPSCGDICNFPPKIAHGHYKQSSSYSFFKEEIIYECDKGYILVGQA
        360       370       380       390       400       410

            120       130       140       150       160       170
B5R    EEKNGNTSWNDTVTCPNAECQPLQLEHGSCQPVKEKYSFGEYMTINCDVGYEVIGASYIS
       . . . . X.....  .: :.  .: .:.      :..:     : .::.:: ::.:.:.   :.
C4BP   KLSCSYSHWSAPAPQCKALCRKPELVNGRLSVDKDQYVEPENVTIQCDSGYGVVGPQSIT
        420       430       440       450       460       470

             180       190       200       210       220
B5R    CTAN-SW-NVIPSCQQKCDMPSLSNGLISGSTFSIGGVIHLSCKSGFTLTGSPSSTCIDG
       :..X .:    .:.:.
C4BP   CSGNRTWYPEVPKCEWETPEGCEQVLTGKRLMQCLPNPEDVKMALEVYKLSLEIEQLELQ
        480       490       500       510       520       530

       230       240
B5R    KWNPVLPICVRTN

C4BP   RDSARQSTLDKEL
        540
```

Fig.20.

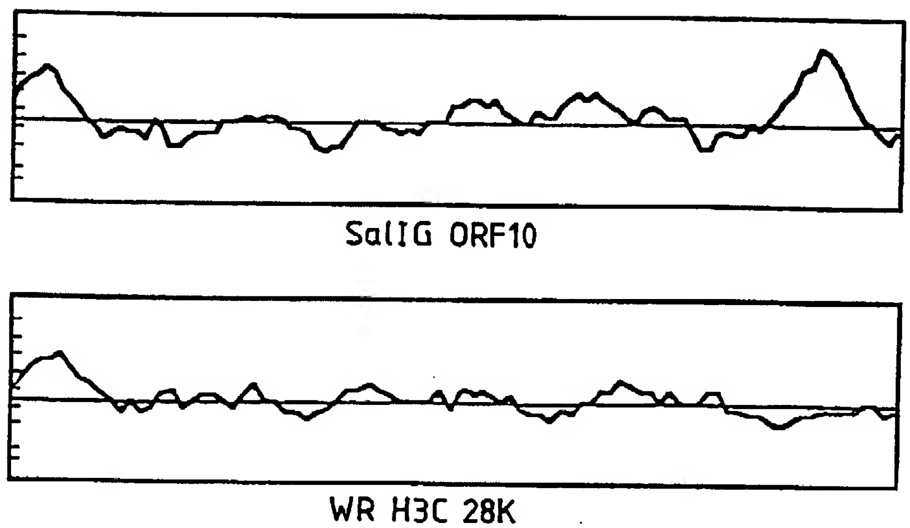


Fig.21.

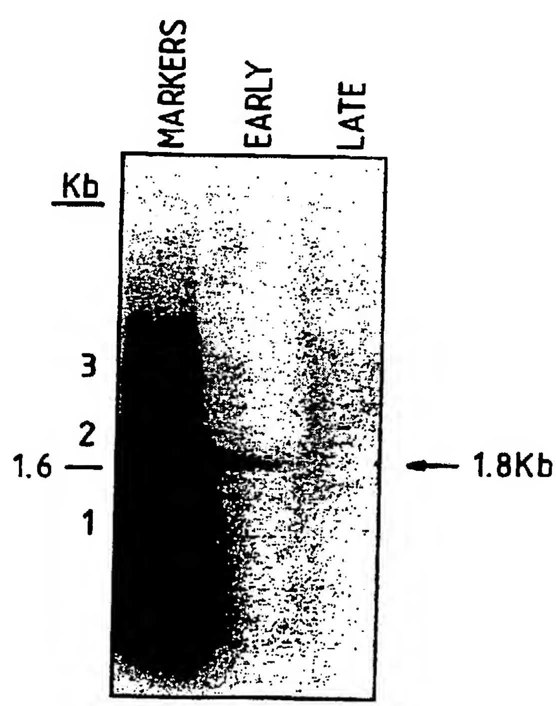


Fig.22.

```
sc MRRLLTGCLLSSARPLKSRLPLL    MSSSLPSSAGKKPKQATLARFFTSMKNKPTEGTPSPKKSSKHMLEERMDNVSGEEEVATKKLKQTAVTHTVAAP
sp  RTVFSQIPRFKQVNQYIRMSTRQSDISNFFISSASHKSEHVEVSQSSSDSKNVDGRSTSEKRKVESVKLVDESKHNNHDDTGTQNVERENNIVSEAKKQ
vv           MTSLREFRKIC       CDIYHASGYK                            EKSKLIRDFITDR        DDKY

sc SSMGSNFSSIPSSAPSSGVADSPQQSQRIVGEVEDALSSNNNDHYSSNIPYSEVCEVFNKIEAISSRLEIIRICSDFFIKIMKQSSKNVIPTTYLFINHL
sp KTLGSSSSSSDAVSSNNDSGASTPIPLPIKEPPLE  SNARNDKLKGHATFAEMVKAFTKIENTSKRLEIIDIMGTYFFGILRDHPSTLLACVYLSINKL
vv                          LIIKILLPGID                      DRIYNMNDK QIIKLYSIIF   KQSQEDML          QDI

sc GPDYEAGLELGLGENLLMKTISEICGKSMSQIKLKYKDIGDLGEIAMGARNVQPTMFKPKPLTVGEVFKNLRAIAKTQGKDSQLKKMKLIKRMLTACKGI
sp GPDY SGLELGIGESIIMKAIGESTGQTLQQIKLSFHKVGDLGLVAQTSRQNQPTMFKPAALTIPFLFDSLKKIAQMSGNQSQNRKIGVIKRLLSSCEGA
vv GYGY                  IGDI        IRTFFKENTEI        RPRDKSILTLE  DVDSFLTTLSSVTK      ESHQIKILTDIASVQTCN

sc EAKFLIRSLESKIRIGLAEKTVLISLSKALLLHDENREDSPDKDVPMDVLESAQQKIRDAFCQVPNYEIVINSCLEHGIMNLDKYCTLRPGIPLKPMLAK
sp EPKYLIRALEGKIRLQLAEKTILVALANATAQYHADKNG   EKLSQQDRIEG EQILRDVYCQLPSYDIITVPHLIEHGLGTLRETCKLTPGIPTKPMLAK
vv DLKCVVMLIDKDIKIKAGPRYVINHISPNAY                          DVFRKSNNLKEIIENASKQNLDSI      SISVMTEINPMLAE
                         *

sc PTKAINEVLDRFQGETFTSEYKYDGERAQVHLINDGTMRIYSRNGENMTERYPEINITDFIQDLDTTKNLILDCEAVAWDKDQGKILPFQVLSTRKRKDV
sp PTKQISEVLNTFDQAAFTCEYKYDGERAQVHFTEDGKFYVFSRNSENMSVRYPDISVSVSKWKKPDARSFILDCEAVGWDRDENKILPFQKLATRKRKDV
vv SCDSVNKAFKHFPSGMEA EVKYDGERVQVH KNNNEFAFFSRNMKPVLSHKVEYLKEYIPKAFKKATSIVLDSE IVLVDEHNVELPFGSLGIHKKKEY

sc ELNDVKVKVCLEAFDILCYNDERLINKSLKERREYLTKVTKVVEGEFQYATQITTNNLDEIQKFLDESVNHSCEGLMVKMLEGPESHYEPSKRSRNWLKL
sp KIGDIKVRACLEAFDILYLNGQPLLETPINERRKILYSMFQPSTGDFTFAKHSDQKSIESIEEFLEESVKDSCEGLMVKMLEGPDSHYEPSKRSRHWLKV
vv KNSNM     CLEVFDCLYFDGFDMTDIPILYERRSFLKDVMVEIPNRIVFSELTNISNESQITDVLDDALTRKLEGLVLKDINGV    YEPGKR  RWLKI

sc KKDYLE  GVGDSIDICVLGAYYGSGKRTGTYGGFLLGCYNQDTGEFETCCKIGTGFSDEMLQLLHDRLTPTIIDGPKATFVFD  SSAEPDVWFEPTTL
sp KKDYLS  GVGDSIDIIVIGAYYGKGKRTSVYGAFILGCYDPDTETVQSICKLGTGFSEEHLETFYNQLKDIVISEKKDFYAHSDVPAHQPDVWFEPKYL
vv KFDYINEGSMADSADIVVLGAYYGKGAKGGIMAVFIMGCYDDESGKWKTVTK CSGHDDNTLRVLQDQLTMVKINKDPKKIPEWLVVNKIYIPDFVVEDP

sc FEVLTADISLSPIYKAGSATF   DKGVSIRFPRFIRIREDKGVEDATSSDQIVELYENQ   SHMQN
sp WEVLAADISLSPVYKAAIGYVQEDKGISIRFPRFIRIREDKSWEDATTSEQVSEFYRSQVAYSQKEKEGSPAAEDY
vv KQSQIWEIS GAEFTSSKSHT   ANGISIRFPRFTRIREDKTWKESTHLNDIVNITKS
```

Fig.23.

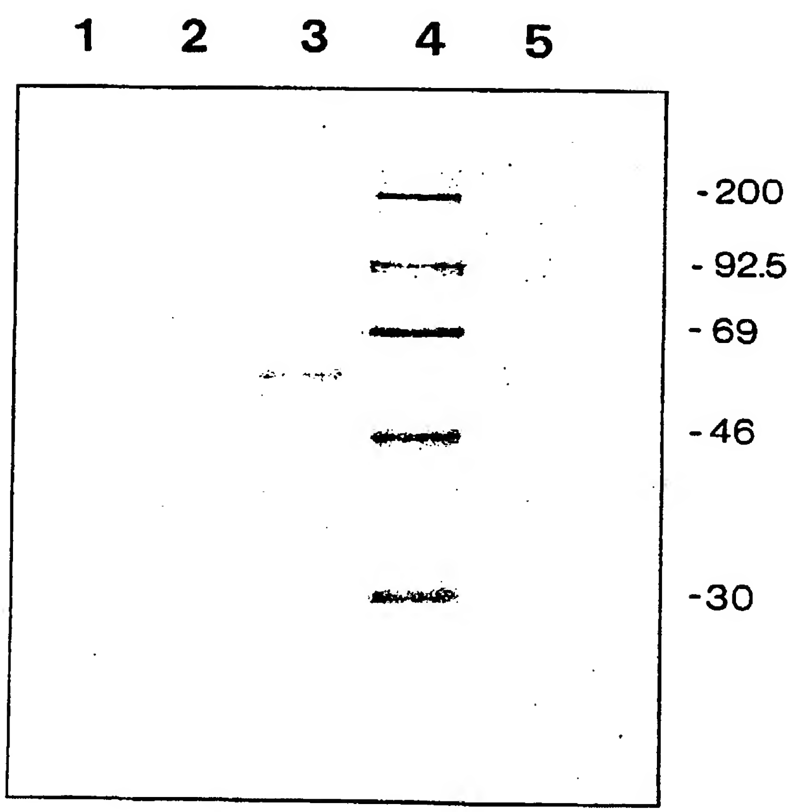


Fig.24.

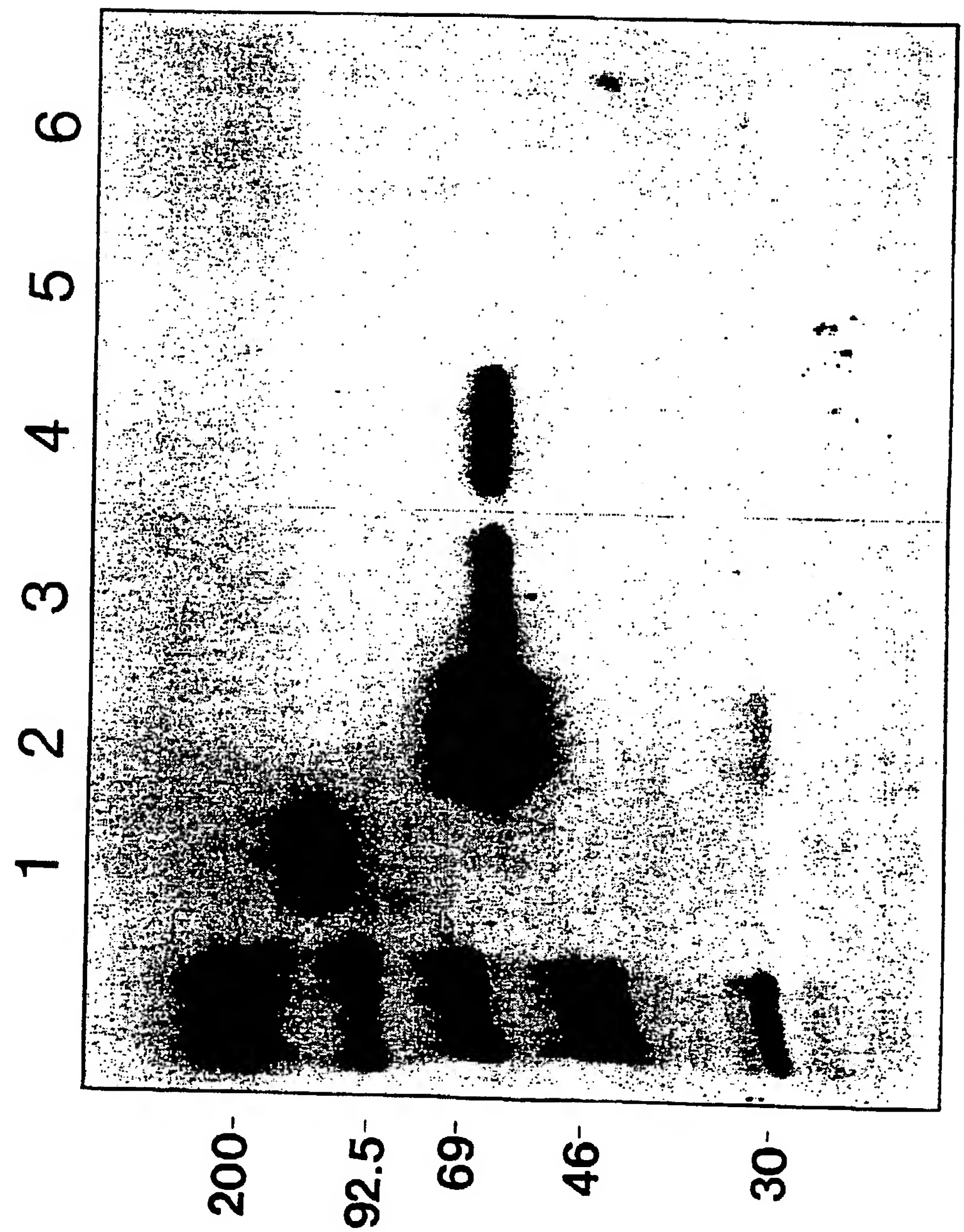


Fig.25.

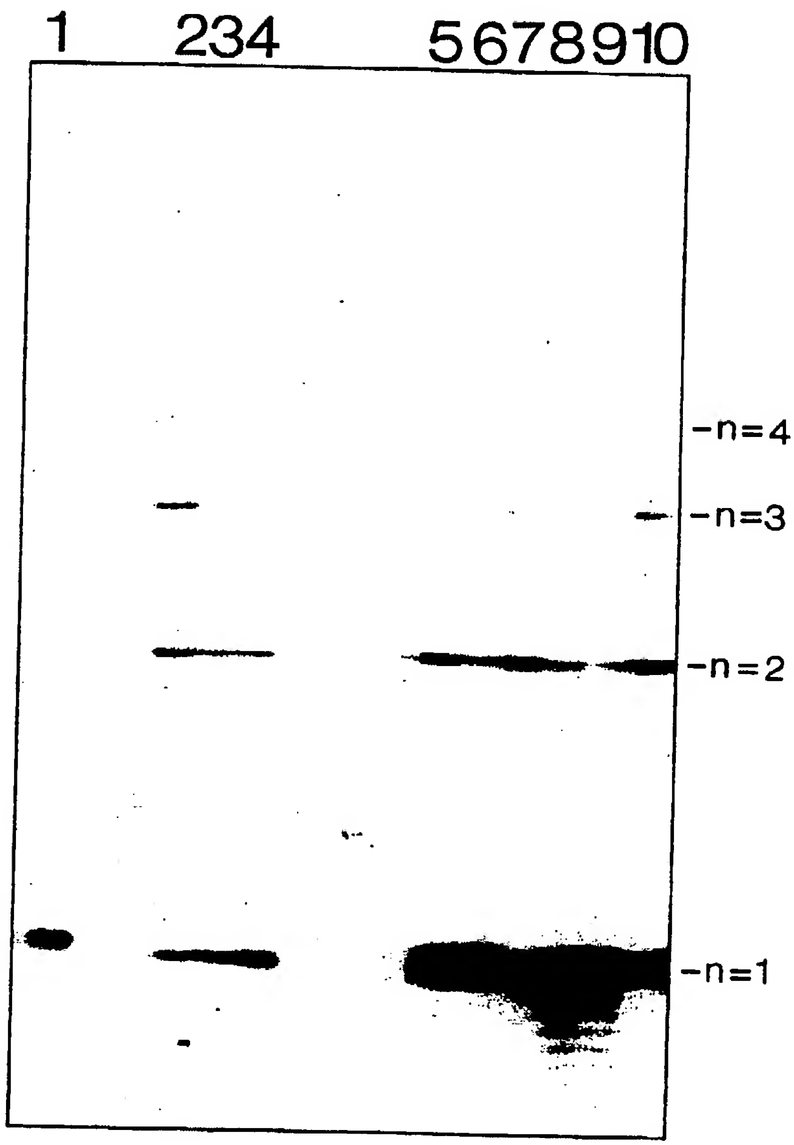


Fig. 26.

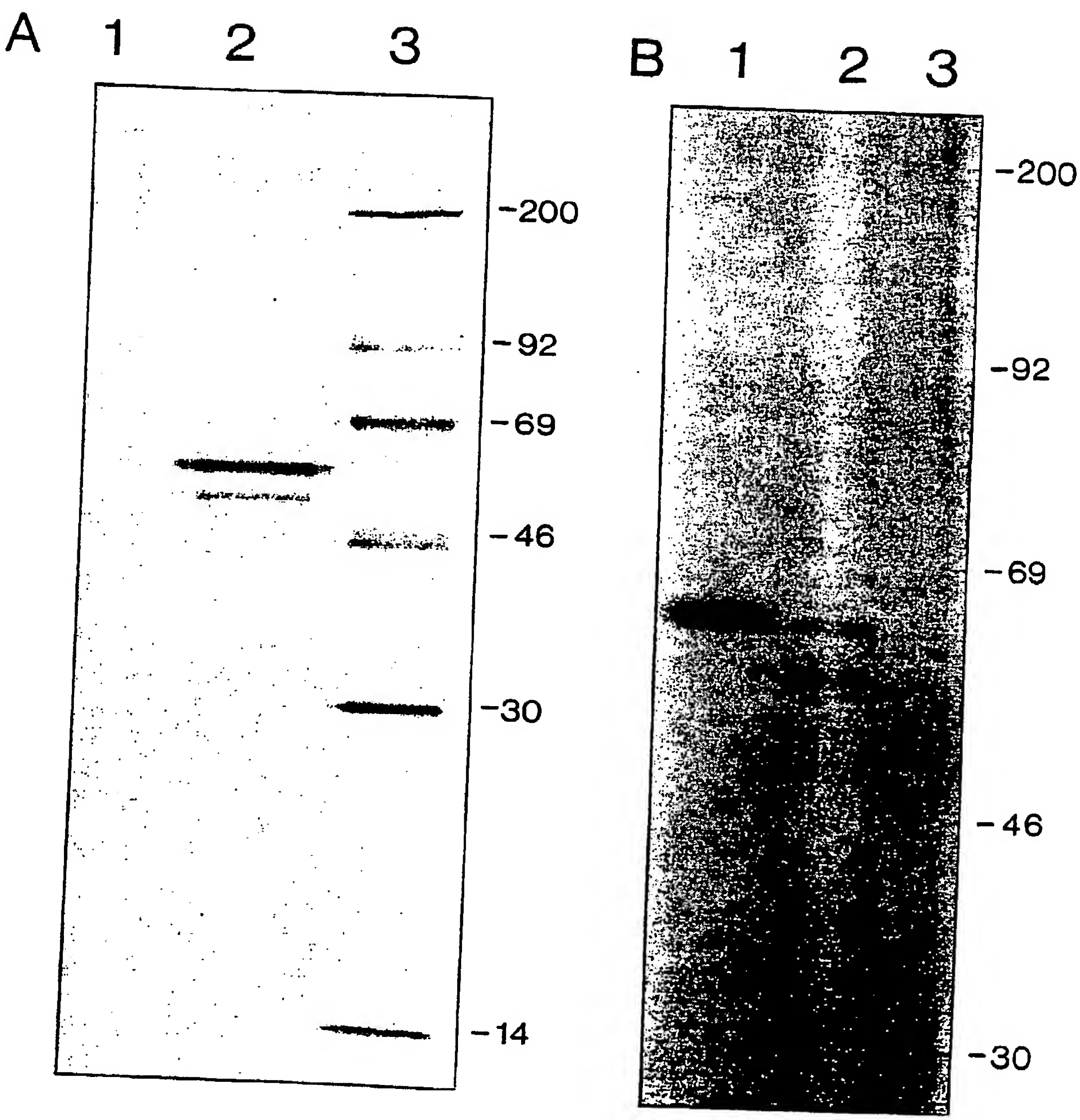
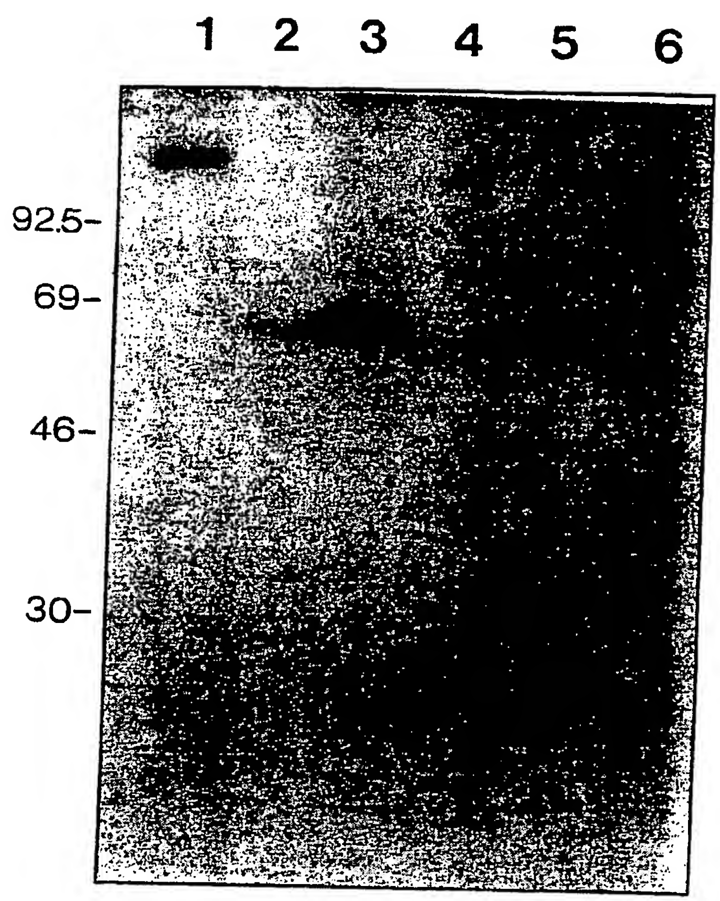


Fig. 27.



26/29

Fig. 28.

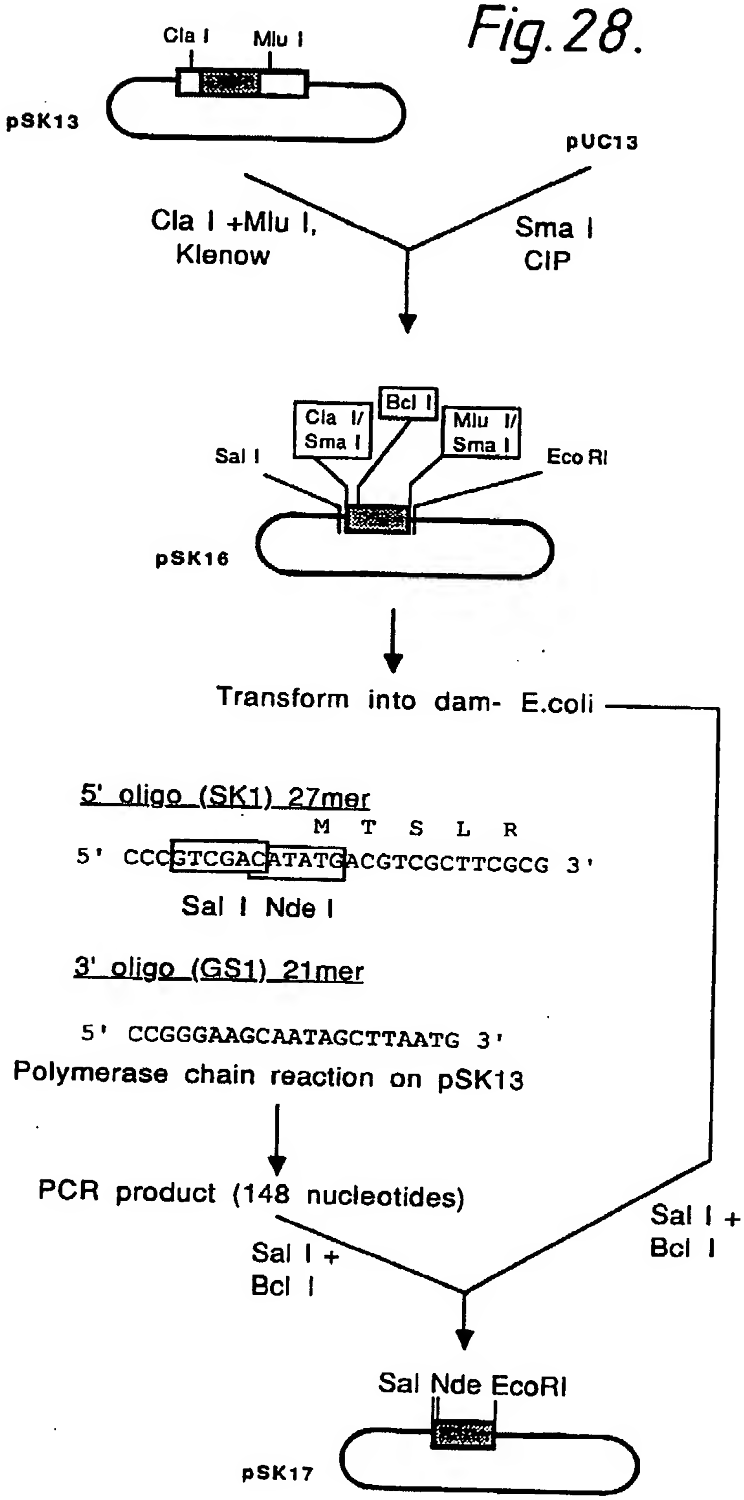

Cla I
Mlu I
pSK13
pUC13
Cla I +Mlu I,
Klenow
Sma I
CIP
Cla I/
Sma I
Bcl I
Mlu I/
Sma I
Sal I
Eco RI
pSK16
Transform into dam- E.coli
5' oligo (SK1) 27mer
M T S L R
5' CCCGTCGACATATGACGTCGCTTCGCG 3'
Sal I Nde I
3' oligo (GS1) 21mer
5' CCGGGAAGCAATAGCTTAATG 3'
Polymerase chain reaction on pSK13
PCR product (148 nucleotides)
Sal I +
Bcl I
Sal I +
Bcl I
Sal Nde EcoRI
pSK17

Fig.29.

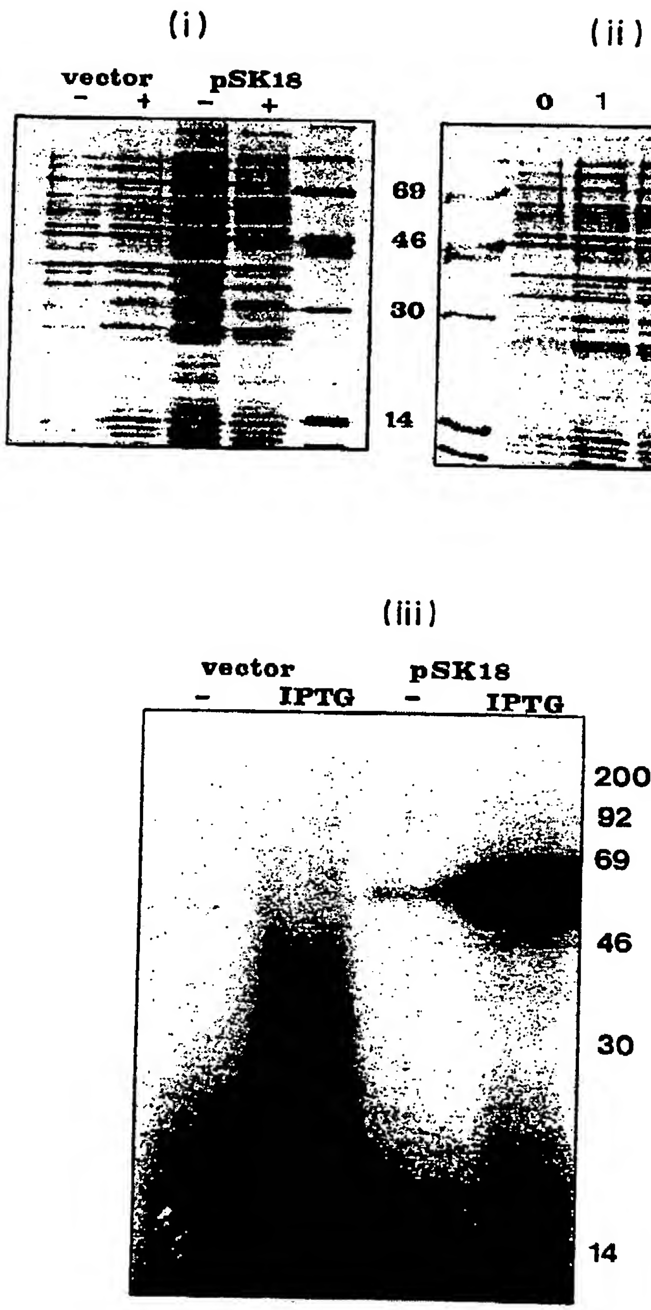

(i)
vector
- +
pSK18
- +
69
46
30
14
(ii)
0 1 2 4
(iii)
vector
- IPTG
pSK18
- IPTG
200
92
69
46
30
14

Fig. 30.

Southern blot of 8 LIGgpt isolates

PROBE = LIG deletion

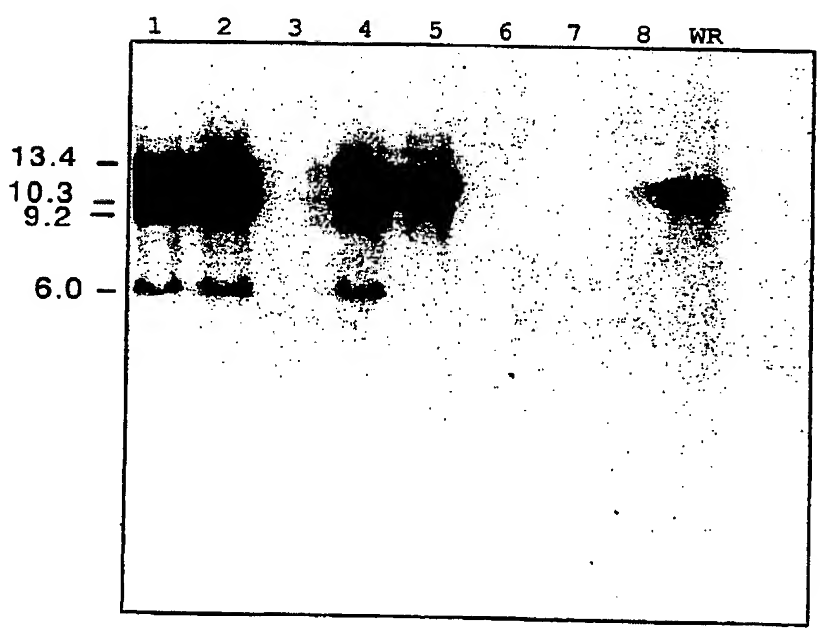
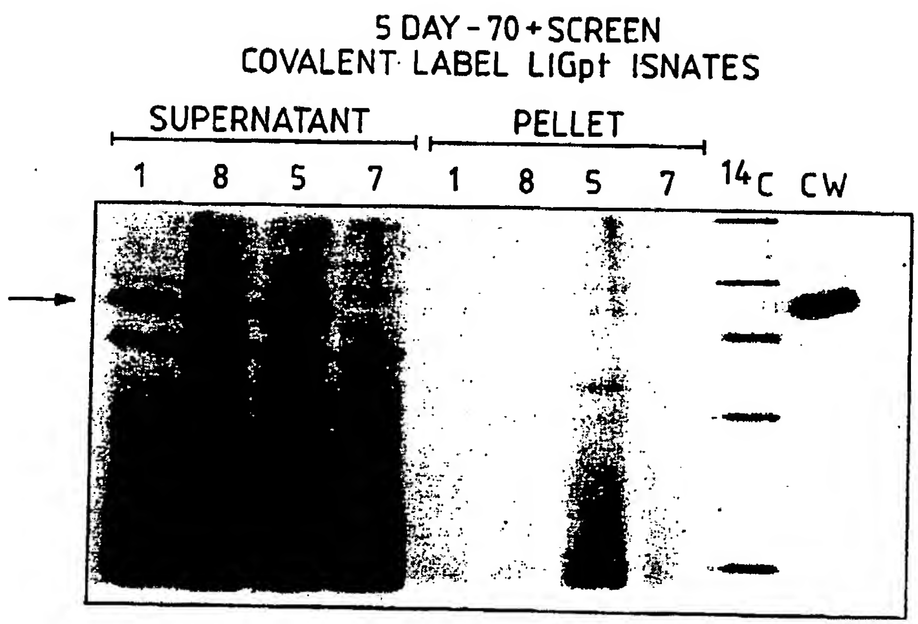


Fig. 31.


5 DAY - 70 + SCREEN
COVALENT LABEL LIGpt ISNATES
SUPERNATANT
PELLET
1 8 5 7 1 8 5 7 14C CW

INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 90/00493

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC[5]: C 12 N 15/86, A 61 K 39/285, C 12 N 15/12, C 07 K 15/04, C 12 N 15/52

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
IPC[5]	C 12 N, A 61 K

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category [*]	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
A	Proc. Natl. Acad. Sci. USA, volume 86, no. 4, February 1989, (Washington, D.C., US), D. Rodriguez et al.: "Highly attenuated vaccinia virus mutants for the generation of safe recombinant viruses", pages 1287-1291 see the whole article --	1-7,13-17
A	Vaccine, volume 5, no. 1, March 1987, (Guildford, Surrey, GB) M. Morita et al.: "Recombinant vaccinia virus LC16m0 or LC16m8 that expresses hepatitis B surface antigen while preserving the attenuation of the parental virus strain", pages 65-70 see the whole article --	1-7,13-17

./.

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
28th June 1990	19.07.90
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	Mme N. KUIPER

Form PCT/ISA/210 (second sheet) (January 1985)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	EP, A, 0255383 (CHIBA) 3 February 1988 see the whole document	1-7,13-17
P,X	Nucleic Acids Research, volume 17, no. 22, 25 November 1989, S.M. Kerr et al.: "Vaccinia virus encodes a polypeptide with DNA ligase activity", pages 9039-9050 see the whole article (cited in the application)	11,12,18

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. GB 9000493

SA 35649

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 17/07/90
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A- 0255383	03-02-88	JP-A- 63036777	17-02-88
		AU-B- 589538	12-10-89

EPO FORM P0479

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82